Multiscale Analysis of Financial Volatility

Bahar Ghezelayagh

School of Economics

University of East Anglia

A thesis submitted for the degree of

Doctor of Philosophy

2013 October

Abstract

This thesis is concerned with the modeling of financial time series data. It introduces to the economics literature a set of techniques for this purpose that are rooted in engineering and physics, but almost unheard of in economics. The key feature of these techniques is that they combine the available information in the time and frequency domains simultaneously, making it possible to enjoy the advantages of both forms of analysis. The thesis is divided into three sections. First, after brifely outlining the Fourier methods, a more flexible technique that allows for the study of time-scale dependent phenomena (motivated from a discussion on Heisenberg's uncertainty principle) namely Wavelet method is defined. A complete account of discrete and continuous wavelet transformations, and wavelet variation is provided and the advantages of wavelet-multiresolution analysis over Fourier methods are demonstrated. In the second section, the statistical properties of financial returns at 1-day, 5-day and 10-day sampling intervals are studied using `S&P500` index for over a decade, and the links between dependence properties of financial returns at lower sampling frequencies are explored. The concepts of temporal aggregation and skip sampling are discussed and the effects of temporal aggregation on long range dependent time series are theoretically outlined and then tested through simulations and empirically via `S&P500`. In the third section, the variation of two years of five-minute `GBP/USD` exchange rate is analysed and the notion of realised variation is explored. The characteristics of the intraday data at different sampling frequencies (5-minute, 30-minute, 60-minute, 10-hour, 1-day, and 5-day) are compared with each other and filtered out from seasonalities using the wavelet multiscaling technique. We find that temporal aggregation does not change the decay rate of autocorrelation functions of long-memory data of certain frequencies, however the level at which the autocorrelation functions

start from move upward for daily data. This thesis adds to the literature by outlining and comparing the effects of aggregation between daily and intra-daily frequencies for the realised variances, which to our knowledge is a first. The effect temporal aggregation has on daily data is different from intra-daily data, and we provide three reasons why this might be. First, at higher frequencies strong periodocities distort the autocorrelation functions which could bring down the decay rate and mask the long memory feature of the data. Second, the choice of realised variance is crucial in this matter and different functions can result in contadictory outcomes. Third, as the order of aggregation increases the decay rate does not depend on the order of the aggregation.

Contents

List of Figures

List of Tables

1

Introduction

1.1 Background

The research on the variation of financial time series has attracted different groups of audience, ranging from hedge fund mangers to macro-economists and econo-physicists. Over the past 3 decades, appealing properties of financial asset returns have been uncovered and have assisted to build appropriate models. At the heart of investment finance, there is the concept of the Efficient Market Hypothesis (EMH). No other concept in finance, has been discussed and scrutinised as much as EMH. The history of EMH spans more than a century [1] and the building blocks of the past 50 years of research depend on whether EMH is accepted, half accepted or rejected.

The Efficient Market Hypothesis (EMH) proposes that future prices cannot be predicted based on past information and one cannot achieve excessive returns using past information. Even if EMH holds, a vast majority of financial economists and risk managers are not interested in predicting the future price but they are rather interested in forecasting the future volatility, and have used for this purpose a variety of parametric and nonparametric econometric models such as generalised autoregressive conditional heteroscedasticity (GARCH) or stochastic regime switching models. Volatility measurement is of particular value to option traders as they are essentially *trading in volatility.*

[1] George Gibson in 1889 in his book *"The Stock Markets of London, Paris and New York"* explains the idea behind efficient markets: *"When shares become publicly known in an open market, the value which they acquire may be regarded as the judgment of the best intelligence concerning them".*

The need for discovering and better understanding the dynamics of financial markets reaches its peak every time the world faces a financial crisis. The recent financial crises over the past 30 years, 1987's Black Monday, 1997's Asian Financial Crisis, early 2000's recession, and the very recent 2008's Global Financial Crisis, have brought further attention to the characteristics of financial time series. These events which are often known as "fat tail" events when do happen have catasrophic consequences which often last a long time and that is why they are some times given the "Black Swans" title[1].

It has been argued that nobody (with a few exceptions such as Janet Yellen and Nouriel Roubini)[2] saw the most recent crisis coming (Annunziata, 2011).

Economists have been widely criticised for ruling out the possibility of sudden variations such as the one on 19 October 1987 when the Dow Jones Industrial Average plunged 22.61% in one day, or the one on 29 September 2008 when the same index saw a 6.98% drop in just a few hours. Assuming that returns are normally distributed, the probability of the 6.98% loss is approximately 10^{-12} which is expected to happen once every billion years. And the probability of the $20-$standard deviation loss that happened on 19 October 1987, is less than one in 10^{50}, which is expected to happen once every 10^{86} years, a number so big that is very hard to imagine [3]. Of course it is now widely accepted that the fundamental cause of the global financial crash was purely human: over-optimism[4]. But under conventional financial theory, the likes of 6.98% and 22.61% daily losses in the history of financial markets were not even supposed to happen. Yet they did.

Perhaps, the assumption of a manageable risk central to all financial models operating at all scales is to be blamed. What is clear is that these models are not capable of justifying a natural concept in financial time series which is the notion of multi-scale features. There was a time when only daily returns were available, and a risk manager was expected to estimate a measure for the volatility (as volatility is an unobserved

[1]The term "Black Swan" is used by Nassim Taleb in capitalized form to refer an event with three attributes: rarity, extreme impact, and retrospective predictability. Please see (Taleb, 2008) and (Taleb et al., 2009) for more details.

[2]Nouriel Roubini is the Yale economics Professor who has been widely known as Dr. Doom as a consequence of predicting the recession. To read an article in New York times about his lecture at International Monetary Fund on 7 September 2006 please see (Mihm, 2008). Janet Yellen is an economist who is now the head of Federal Reserve. The Fed's December 2007 transcripts reveal that she thought the possibilities of a credit crunch developing seem too real.

[3]Remember that the universe is only 10^{10} years old.

[4]Interested readers are referred to: (Stanton et al., 2010) and (Annunziata, 2011)

variable) using daily data. However nowadays intraday data has become widely available and we are now bound to utilise the more dense observations to maximise the efficiency of our estimations. It would be possible to have an even more precise estimate of daily volatility and to virtually treat daily volatility as an observed variable. With data being available at different frequencies what has become clear is that, financial data contain several structures, each occurring on a different time scale. Clearly, a new model needs to be flexible enough to justify abrupt changes and should capture the multi-scale behaviour of the data and should also be compatible with the statistical properties of high frequency data.

The study of high-frequency data has shed lights into the field of finance. The availability of such data has made the unobservable process of volatility, observable. Modelling the second moment of returns has suddenly changed from parametric measures to simple nonparametric models.

1.2 Time, Frequency, Time-Frequency, and Time-Scale

One of the motivations behind the analysis of a time series in the frequency domain is to find out any periodic features of the data. If there is only one cycle in the time series, then the autocorrelation function reveals this cycle. However, if a series contains several cycles, the autocorrelation function is not capable of separating out such periodicities. The ideal situation is to be able to determine how many cycles exist in the data and how important each cycle is with respect to each other. The frequency domain approach has the power to reveal the information from the data that is not readily available from the data in the time domain.

A time series is a variable that changes successively in time. Electromagnetic waves, speech or audio signals, temperature readings, or stock prices are all examples of time varying signals. One way to represent the signal is to show how the amplitude of a signal changes over time (which is the common way in economics). But it is not the only way. In fact, in many other fields (such as in physics, engineering, astrophysics, geophysics, psychology, language, biology and music) the frequency representation of a signal is the one that matters most. In these areas statistical analysis in the time domain do not play a major role but it is rather in the frequency domain that the characteristics of a signal are analysed.

Region	Frequency
Radio	$< 3 \times 10^9$
Microwave	$3 \times 10^9 - 3 \times 10^{12}$
Infrared	$3 \times 10^{12} - 4.3 \times 10^{14}$
Visible (Red-Violet)	$4.3 \times 10^{14} - 7.5 \times 10^{14}$
Ultraviolet	$7.5 \times 10^{14} - 3 \times 10^{17}$
$X-$ rays	$3 \times 10^{17} - 3 \times 10^{19}$
Gamma rays	$> 3 \times 10^{19}$

Table 1.1: Electromagnetic waves

To realise how different the frequency domain environment can be, consider visible light as an example. Visible light is an electromagnetic wave and its frequency can vary between 4.3×10^{14} to 7.5×10^{14} and it is the *frequency* of the wave that determines the colour of the wave (Someda, 2006). The frequencies inbetween form the colours in the rainbow. Table 1.1 summarises the frequencies for visible and invisible waves.

In the time domain a variable Y_t at time t is analysed in models in which the focus is on the properties of the variable dependent on time t, such as covariance between Y_r and Y_s at distinct dates r and s. In the frequency domain variable Y_t is analysed as a weighted sum of periodic functions of the form $cos(\nu t)$ and $sin(\nu t)$:

$$Y_t = \mu + \int_0^{\pi} \alpha(\nu) \cdot cos(\nu t)\, \mathrm{d}\nu + \int_0^{\pi} \beta(\nu) \cdot sin(\nu t)\, \mathrm{d}\nu \tag{1.1}$$

The goal in the frequency domain is to find out how important cycles of different frequencies are. This thesis is not trying to explain the techniques in the frequency domain but it is rather suggesting that analysing a variable in both frequency and time domains could lead to insightful findings. This thesis is proposing that the statistical analysis of time series in economics and finance would be explored *better* in both two domains.

Fourier transformations are the most common way to transform the information of a variable from time domain to frequency domain, but they are not widely used in standard econometrics. If the variable is observed originally in the time domain then Fourier transform, summarises the information of the time series in an alternative way rather than in time. It summarises the content of the variable as a function of frequency and does not involve with any time information. Short Time Fourier Transform (here-

after abbreviated to STFT) was developed for bringing a balance between time and frequency by sliding a window across the time series and taking the Fourier transform of the windowed series. Perhaps one of the reasons Fourier transforms are not very popular among economists is because Fourier transforms look only for sinusoids globally and therefore are not suitable for evolving behavior of stock markets, as complete sinusoid patterns are rarely encountered in empirical finance. The time information is lost completely in Fourier analysis, whereas, time information is very critical in economics and finance. STFT in only applicable for *covariance stationary* processes and once the window function is chosen, the frequency and time resolution of the window are fixed for all frequency bands and all times. So for a signal observed throughout a length of time, which has different frequency components which might appear, disappear, or re-appear at different times, STFT might not be the most suitable technique to work with.

This thesis provides a solution to overcome the shortcomings of STFT by using Wavelet Transformations. Wavelet analysis is a non-parametric method that allows for the study of time-scale dependent phenomena. Wavelet Transformations are akin to Fourier Transforms but do not lose the time dimension in the transformation. The wavelet transform differs from the STFT by using an entirely different set of basis functions (not sinusoids) which adaptively partition the time-frequency plane to better capture the range of low-frequency to high-frequency events.

The power of wavelets is their ability to analyze (decompose) features of a signal which vary over both time and scale. A natural concept in financial time series is the notion of multi-scale features. That is, an observed time series may contain several structures, each occurring on a different time scale. For instance, stock market consists of multiple layers of investment horizons (time-scales) varying from milliseconds to years. Wavelet techniques possess an inherent ability to decompose time series into several sub-series which may be associated with a particular time scale. In other words, wavelet methods present a *lens* to the researcher, which can be used to zoom in on the details and draw an overall picture of a time series at the same time. Unlike Fourier transforms, they have the capability to decompose processes on different time scales, but still preserve time localization.

The main motivation in this thesis is to uncover the characteristics of volatility measures at different scales ranging from intraday to more than a day sampling in-

tervals. The activities in financial markets involve dealers in different geographical locations who have different time horizons, home currencies, information access, transaction costs, and risk attitudes. The time horizons vary from computer traders trading at nanoseconds to intraday dealers, who close their positions every evening, to long term investors and central banks who make decisions depending on much longer horizons (months, quarters or even years). In such a complex structure, market participants follow different strategies to reach their goals. The development of computer technology, has eased the paths for academic researchers to have access to intraday data. With more information available, there has been a rapidly growing body of empirical studies on the behavior of the intra-daily volatility measures.

For decades historical volatility has been an unobserved variable but as with the advancement in technologies intraday data have become available, and time horizons such as secondly and minutely can now be used, volatility could "virtually" be treated as an observed variable. Using intra-day observations for estimating volatility, has replaced "historical" volatility with "realised" volatility. The use of intra-day data and notion of realised variance has given birth to the ongoing new research course of volatility modeling using intra-day data. This thesis utilises Wavelet methodologies and apply them on intraday data to further uncover the characteristics of the data.

1.3 Plan of the thesis

This thesis is divided into four parts. The second chapter of the thesis after the introduction starts by investigating the drawbacks of STFT. In doing so, the concepts of time and frequency localisation with regards to Heisenberg's Uncertainty Principle are explored: it is impossible to achieve an optimal resolution simultaneously in the time and frequency domains. This obvious fact then motivates the need for a more flexible environment. Wavelet transforms are defined as mathematical transformations that intelligently overcome the Heisenberg's principle without violating the theorem. Continuous and discrete wavelet functions, filters, and coefficients are then defined. Different examples of wavelet family are provided which are used for comparisons against Fourier transforms. The perception of multiresolution analysis using wavelet decompositions is expounded, which leads into the definition of wavelet variance and covariance.

The third chapter used 10 years of daily observations of S&P500 and starts with describing the data and giving the definitions of the asset returns. The dependence properties of financial returns are discussed and concepts such as long-memory processes, Autoregressive Fractional Moving Average (ARFIMA) are explained in details. Different methods for testing and estimating ARFIMA models are expounded, namely the R/S test statistic, the GPH estimator, and the Wavelet Maximum Likelihood Estimator. Section 3.5 explores other properties of financial returns and the famous question "What is the relevant horizon for risk management?" is proposed and answered. This chapter then continues by exploring concepts of skip sampling and temporal aggregation followed by identifying the effects of temporal aggregation on ARFIMA(0,d,0) theoretically and through simulations. This section ends by analysing the $S\&P500$ daily, 5-day and 10-day returns. Some of the proofs are provided in the appendix.

The third chapter starts by studying the properties of intraday data and it used the tick-by-tick and also 5-minute GBP/USD exchange rates for different periods of time, one month and two years respectively. The concept of realised variance and different ways of approaching the estimator are discussed. The effects of temporal aggregation are dealt with and the chapter ends with considering two different scenarios for forecasting the realised variance.

The fourth chapter is the conclusion which brings the results of chapters two and three together and gives suggestions for future research.

This thesis has used a selection of widely available software: **R** version 3.0.1 for some of the estimations and diagrams. Whenever a package is used, corresponding authors are referenced. **R** is an object oriented General Public License software which is written in **C**, Fortran and **R** languages[1]. **Matlab** version $2010b$ (mainly the Wavelet toolbox for showing pictures of different wavelet functions)[2], and **Stata** version 13.1 [3] for some of the statistics and diagrams in chapter four. Most of the R codes used for plotting figures and estimating models are available in appendix.

[1]Available at: http://www.r-project.org/
[2]Available at: http://www.mathworks.co.uk/products/wavelet/
[3]http://www.stata.com/

2

Wavelet Methods for Multi-Resolution Analysis

2.1 Introduction

Wavelets are a relatively new technique for representing the levels of details present in the function. Wavelets are mathematical tools for hierarchically decomposing functions which could resemble anything; time series, images, or even curves (see (Stollnitz et al., 1996) for other applications). This decomposition will make it possible to describe an object in terms of a coarse shape and a range of details. The birth of the formal subject of wavelets in time series analysis and digital signal processing goes back to 1980s, however it has only been over the last decade that wavelets have started to find their ways into economics. From a more general perspective, wavelet analysis has rather completed the existing analysing techniques in some areas such as in spectral analysis, but in some areas it has been capable of shedding lights on matters where little progress has been made such as in de-noising.

Most of the existing literature of wavelets in time series analysis require extensive knowledge of Fourier and Windowed Fourier Transforms (with the exceptions of (Ramsey, 1999),(Ramsey, 2002), and (Rua, 2012)). This chapter of the thesis is trying to familiarise the readers with wavelet techniques without going through all the extensive mathematical proofs. References are given for interested readers to seek the full proofs. However, before defining wavelet techniques, one needs to go on a quick journey

from time domain to the frequency domain, and then from time-frequency domain to time-scale domain. This will help to understand the intuition behind the reasons why wavelet approach is the preferred (and complementary) method compared to existing methods for analysing financial time series in domains other than the time domain. Wherever needed, examples have been given to ease the flow of the discussions.

This chapter starts with giving brief definitions of spectral analysis and Fourier transform. After familirising the reader with spectral tools, the chapter extrapolates the shortcomings of (Short Time) Fourier Transforms, defines two dimensional time-frequency plane and discusses the limitations of time-frequency space using Heisenberg uncertainty principle. By building up the motivation for a new platform in which time series can be analysed more precisely, wavelet functions, continuous and discrete wavelet transforms are defined. Examples of the wavelet family are provided, the concept of multiresolution analysis is explained and wavelet variance are defined.

2.2 Spectral Analysis

A time series is a set of observations made sequentially in time and it is a process that is varying in time. Often a graph of its value plotted against time is referred to as the variable itself. A time domain analysis is aimed to reveal the sequence of events within the data. That is, the value of a variable (Y_t) at time t is described in terms of a sequence of innovations $[\epsilon_t]_{t=-\infty}^{t=+\infty}$ in models of similar form:

$$Y_t = \mu + \sum_{i=0}^{i=\infty} \Psi_i \epsilon_{t-i} \tag{2.1}$$

where the focus is on analysing the properties of $[Y_t]_{t=-\infty}^{t=+\infty}$ in the time domain and how different functions of the return series of Y_t are inter-connected. However, this is not the only way to describe, analyse and study a time series, signal, variable, or a function. There are two approaches in time series analysis: time domain analysis and frequency domain analysis. A frequency domain analysis describes the data in terms of sinusoidal functions to reveal its component sequences whenever they exist in separate frequency bands. Concepts that emerge frequently in frequency domain are: phase(a), amplitude(A), frequency(ν), and bandwidth (see figure (2.1)). Phase is the location of

the peak and the trough of the oscillation, amplitude shows the size and power of the signal.

$$Y_t = Asin(a + \nu t) \tag{2.2}$$

Figure (2.1) shows a simple sinusoidal function for $\nu = 1, a = 0, A = 1, t = 1, ..., 20$.

Figure 2.1: The distance between the two dotted vertical lines (blue line and the green line) is the bandwidth.The R codes for plotting the graph is given in appendix.

However, time series in real world do not always look like figure (2.1) and possibly have several sinusoidal components, with different amplitudes, different phases, and different frequencies. In the frequency domain variable Y_t is analysed as a weighted sum of periodic functions of the form $cos(\nu t)$ and $sin(\nu t)$:

$$Y_t = \mu + \int_0^\pi \alpha(\nu_1) \cdot cos(\nu_1 t)\, \mathrm{d}\nu_1 + \int_0^\pi \beta(\nu_2) \cdot sin(\nu_2 t)\, \mathrm{d}\nu_2 \tag{2.3}$$

One of the similarities in between the time and the frequency domains is that, the analyses are based on the assumption of covariance stationarity. The process Y_t is said to be covariance stationary if neither the mean μ_t nor the autocovariances κ_t depend on time t. That is, $E(Y_t) = \mu$ for all t and $cov(Y_t, Y_{t+\tau}) = \kappa(\tau)$ for all t and for all τ.

Standard time series text books that have included spectral analysis are: (Priestley, 1981), (Harvey, 1991), (Hamilton, 1994) (799 pages, with only 61 pages on spectral analysis and the Kalman Filter), (Koopmans, 1995), (Bloomfield, 2000), (Osborn and Ghysels, 2001). In economics and finance, the interest in studying business cycles in 1940s led to a new line of research dealing with seasonal fluctuations of macro variables such as GDP (Burns and Mitchell, 1946). However, research on seasonal fluctuations seemed not to have received a reasonable interest and economists were relatively reluctant to use a new approach (Miron, 1996). In finance, analysis in frequency domain didn't catch on with the econometric analysis in the time domain, even though there have been several attempts towards bringing the two empirical analyses together. On one hand, *... the time and frequency domain approaches should not be considered as competitors and they should rather complete each other*(Harvey, 1975). On the other hand, *... applications of spectral analysis to econometric time series will be less useful*

... periodicity of most econometric time series, such as business cycles or quarterly effects, can easily be identified, spectral analysis in these areas will be of only marginal value (Chan, 2010). It is worth noting that, "time series" might not necessarily mean that one is limited in using methods in the time domain but rather it means the original series is a variable that changes successively in time, hence *time series*.

The most common transformation that translates the information of a time series from the time domain to the frequency domain is Fourier transform. During the past two decades there has been a gradual increase in the application of Fourier methods in standard econometric text books [for example: (Wang, 2009) and (Belsley and Kontoghiorghes, 2009)], and analysis in frequency domain is now regarded as an alternative approach that helps capturing cyclical behaviour of the data series.

This thesis shall not explain Fourier methods in details but since spectral analysis using Fourier series act as prerequisites for the rest of the thesis, this chapter starts by Fourier transforms.

2.2.1 Fourier Transform

$f(x)$ is said to have finite energy if:

$$\int_{-\infty}^{\infty} |f(x)|^2 dx < \infty \tag{2.4}$$

And if $f \in L^2(\mathbb{R})$, then its Fourier transform $\hat{f}$ is defined as:

$$\hat{f}(\nu) = \int_{-\infty}^{\infty} f(x) e^{-i\nu x} dx, \nu \in \mathbb{R} \tag{2.5}$$

where ν is the frequency measured in radians per unit time, i is the square root of -1, and $\hat{f}(\nu)$ is the function of frequency. If scalar product is defined as:

$$\langle f(x), g(x) \rangle = \int_{-\infty}^{\infty} \overline{f(x)} g(x) dx \tag{2.6}$$

Then equation 2.5 can be re-written as:

$$\hat{f}(\nu) = \langle e^{i\nu x}, f(x) \rangle \tag{2.7}$$

Fourier transform preserves the energy in $f(x)$. That is, if the inverse Fourier transform is defined as:

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\nu) e^{i\nu x} d\nu = \langle e^{-i\nu x}, \hat{f}(\nu) \rangle \tag{2.8}$$

Then:

$$\int_{-\infty}^{\infty} |\hat{f}(\nu)|^2 d\nu = \int_{-\infty}^{\infty} |f(x)|^2 dx \tag{2.9}$$

Equation (2.9) is a very important property of Fourier transforms and is known as Parseval's identity and is often stated more generally as:

$$\langle f(x), g(x) \rangle = \langle \hat{f}(\nu), \hat{g}(\nu) \rangle \tag{2.10}$$

It basically states that the integral of the square of the Fourier transform of a function is equal to the integral of the square of the function itself.

Technically, the right-hand side of (2.5) defines a function $\hat{f}$ only when the original f is absolutely integrable. In practice, one uses:

$$\hat{f}_n(\nu) = \int_{-n}^{n} f(x) e^{-i\nu x} dx, \nu \in \mathbb{R} \tag{2.11}$$

Under certain conditions,

$$\lim_{n\to\infty} ||\hat{f} - \hat{f}_n||^2 = \lim_{n\to\infty} \int_{-\infty}^{\infty} |(\hat{f}(\nu) - \hat{f}_n(\nu)|^2 d\nu = 0 \tag{2.12}$$

Equation (2.12) is the mathematical definition of the Fourier transform. However, the data used in this thesis (and mostly in finance and economics) are inherently discrete signals of finite length and energy. Hence, it is more practical to use:

$$\hat{f}(\nu) = \frac{1}{2\pi} \sum_{-n}^{n} f(\Delta x) e^{-i\nu\Delta x}, -\pi \leq \nu \leq \pi \tag{2.13}$$

Equation (2.13) can be used to provide the link between time domain analysis and frequency domain analysis. If κ_τ, the autocovariance functions is defined as: $\kappa_\tau = cov(X_t, X_{t+k}) = E[(X_t - \mu_t)(X_{t+k} - \mu_{t+k})]$ and if replaced as $f(\Delta x)$ in 2.13, then:

$$\hat{f}(\nu) = \frac{1}{2\pi} \sum_{-n}^{n} \kappa_{-i\nu\tau, -\pi \leq \nu \leq \pi} (2.14)$$

where, ν is frequency measured in radians per unit time, and $f(\nu)$ is a function of ν called *power spectrum*. Equation (2.14) is used in this thesis for calculating power spectrum using Fourier transforms. Since the data used in this thesis (and generally in finance) do not have any imaginary parts and are all real observations[1], using:

$$e^{-i\nu\tau} = \cos(\nu\tau) - i\sin(\nu\tau) \tag{2.15}$$

equation 2.14 can be re-written as:

$$\hat{f}(\nu) = \frac{1}{2\pi}[\Gamma(0) + 2\sum_{\tau=1}^{\infty} \Gamma(\tau)\cos(\nu\tau)], 0 \leq \nu \leq \pi \tag{2.16}$$

which is basically the Fourier transform of the autocovariance function (κ_τ). An interesting case is the power spectrum of a white noise. $Y_t = \epsilon_t$, with $E(\epsilon_r, \epsilon_s) = \sigma^2$ for $r \neq s$ then $\gamma(\tau) = \sigma^2$ and it follows from equation (2.14) that $\hat{f}(\nu)$ for a white noise is a flat line ($\frac{1}{2\pi}$). It is worth mentioning that white noise is called *white* noise because it contains equal power across frequency axis which is parallel with white light which is an equally weighted average of all colours.

Instead of plotting $_\tau$ (a function depending on time) against τ, one can plot $\hat{f}(\nu)$ (a function depending on frequency) defined in equation (2.16) against ν. The autocovariance function (κ_τ) and spectral density are very close, since:

$$\kappa_\tau = \int_{-\pi}^{\pi} f(\nu) e^{i\nu\tau} d\nu \tag{2.17}$$

which results in:

$$(0) = \int_{-\pi}^{\pi} f(\nu) e^{i\nu 0} d\nu = \int_{-\pi}^{\pi} f(\nu) d\nu \tag{2.18}$$

which means that the area under the spectral density function of X_t between $-\pi$ and π gives the variance of X_T. Worth remembering that the long-run variance is

[1] A complex number such as z, is written as $z = a + ib$, where a and b are both real numbers. a is known as the real part of z and b is known as the imaginary part of z. The conjugate of z which is often shown by $\overline{z}$ is: $\overline{z} = a - ib$

equivalent to the spectral density at the zero frequency.

To observe how Fourier transforms can be useful in adding to the results of time domain analysis, two examples are given. Two simulated series: y_t and w_t which are AR(2) and AR(3) processes respectively, are defined as follows:

$$y_t = 0.5y_{t-1} - 0.7y_{t-2} + \epsilon_t, t = 1, ..., 512 \tag{2.19}$$

and

$$w_t = 0.6w_{t-1} - 0.5w_{t-2} - 0.2w_{t-3} + \epsilon_t, t = 1, ..., 512 \tag{2.20}$$

In order to look at the roots of the *characteristic* function, equations (2.19) and (2.20) can be written in lag polynomial notations:

$$(1 - 0.5L + 0.7L^2)y_t = \epsilon_t, t = 1, ..., 512 \tag{2.21}$$

and

$$(1 - 0.6L + 0.5L^2 + 0.2L^3)w_t = \epsilon_t, t = 1, ..., 512 \tag{2.22}$$

y_t has roots, $0.357 \pm 1.141i$ and w_t has roots $0.592 \pm 1.004i$ and -3.683.

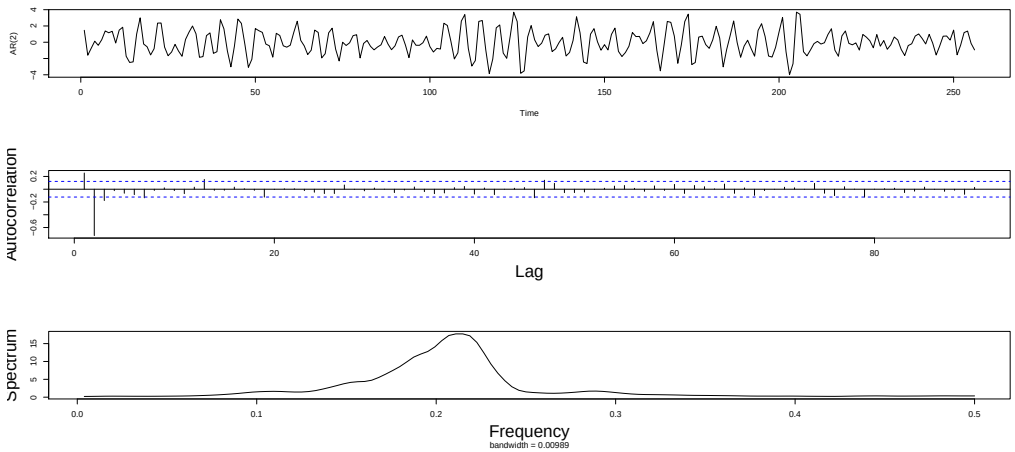

Figure 2.2: From top to bottom: Plots of simulated AR(2), y_t (t=512) against time [2.19], partial autocorrelation functions against lag, and power spectrum against frequency.

Figure (2.2) provides full information about y_t. From top, a plot of the variable against time, sample autocorrelation functions up to 90 lags, and power spectrum against frequency. Figure (2.3) shows the same for w_t. The Fourier transformation has transformed the information of y_t and w_t from the time domain to the frequency domain which have been plotted at the bottom of both figures against frequency. This extra information added, helps to identify any cyclical behaviour in the time series.

What this thesis is concerned with is how the information from the time domain has been transferred to the frequency domain. To illustrate this: figure (2.5) shows how originally information of the signal was observed in the first place and figure (2.6) shows how Fourier transformation has translated the information.

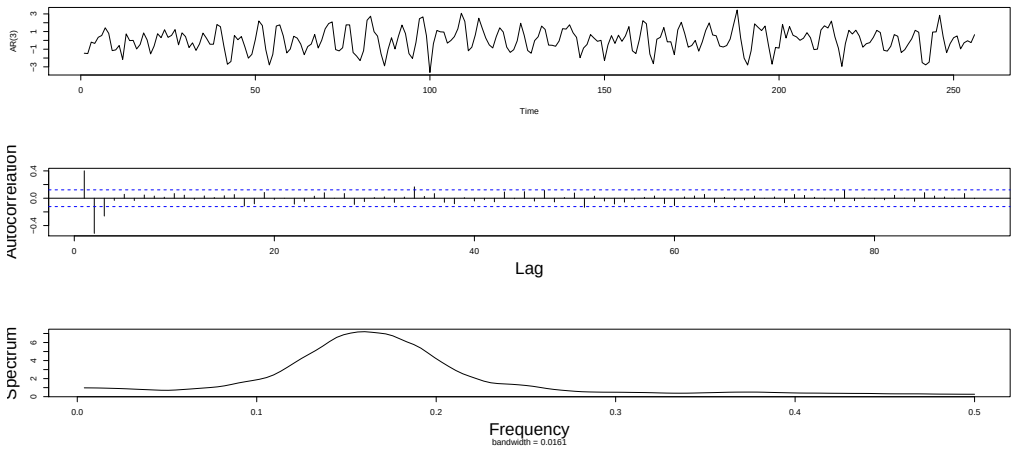

Figure 2.3: From top to bottom: Plots of simulated AR(3), w_t (t=512) against time [2.20], partial autocorrelation functions against lag, and power spectrum against frequency.

From figure (2.6), one cannot extract any information about the content of the signal in the time domain and from figure (2.5), the frequency content of the signal is not readily available. That is not surprising since, on one hand all the formulae used to analyse time series in the time domain do not have a place for frequency (ν) and on the other hand, equation (2.16) gives information about the frequency content of the time series and does not give any information regarding the location of these frequencies in the time domain. But could the fact that equation (2.16) does not provide any information regarding the time dimension of the signal, be a problem? Yes, we explain with an example.

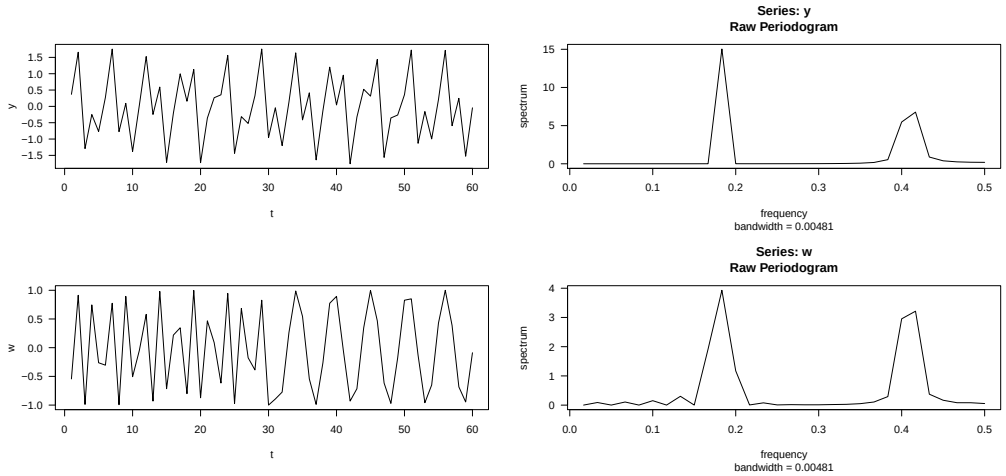

Figure 2.4: The first signal (top) consists of superposition of two frequencies $sin(10t)$ and $sin(20t)$, and the second (bottom) consists of the same two frequencies each applied separately over half of the signal duration.

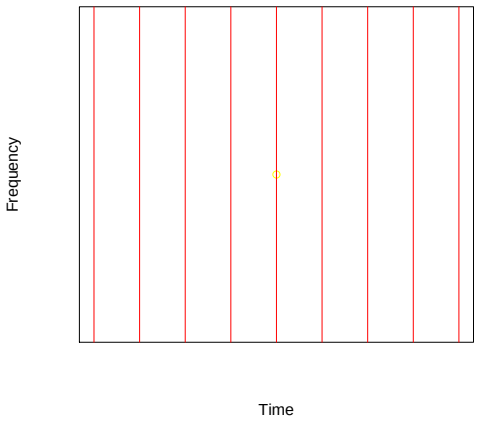

Figure 2.5: The time domain representation of the observed time series. Perfect time resolution and no frequency resolution!

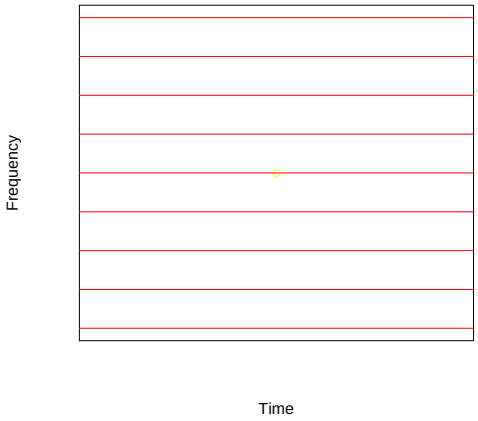

Figure 2.6: The frequency domain after computing the Fourier transform. Perfect frequency resolution and no time resolution!

For example, figure (2.4) shows two time series: the first one consisting of two frequencies ($sin10t$ and $sin20t$) for the entire duration of the signal ($t = 60$), and the second one consisting of the same frequencies ($sin10t$ and $sin20t$), but each is applied separately for half of the signal duration: for the first half ($sin10t$) and for the second half ($sin20t$)[1]. As can be seen, the spectrum is quite incapable of realising the difference between the two signals, even though they are totally two different series in the time domain.

2.3 The Short Time Fourier Transform (STFT), or Gabor Transform

As discussed Fourier transforms exhibit the spectral content of the time series in frequency domain, but does not give any information regarding where in time those spectral components actually appear, as seen in figure (2.4). This is simply because of the fact that the sine and cosine functions are not local but global functions and cannot cope with disappearing or re-appearing a feature of the data. Hence, Fourier transform on its own is not a suitable technique for time evolving time series such as the ones in

[1]please see appendix for the R codes to generate figure (2.4)

high frequency data in finance. Fourier transforms are useful when one is interested in knowing what spectral components have appeared in the time series but not interested when "in time" (the example in the introduction), which is not the case in financial economics.

What is often done to overcome this problem is to divide the whole time interval into several shorter time intervals. Then take the Fourier transform for each interval, which is actually what most of the engineers and physicists do. Because the time interval is divided into shorter time intervals, this method is called Short Time Fourier Transform (STFT) or Gabor transform[1]. If $f \in L^2(\mathbb{R})$, then its Fourier transform $\hat{f}$ was defined as:

$$\hat{f}(\nu) = \int_{-\infty}^{\infty} f(x)e^{-i\nu x}dx, \nu \in \mathbb{R} \tag{2.23}$$

So what needs to be explained is how to modify the Fourier transform which can be achieved by multiplying $f(x)$ by a window function φ:

Suppose that φ also $\in L^2(\mathbb{R})$, for any $\nu \in \mathbb{R}$ and $b \in \mathbb{R}$:

$$\Phi_b(f)(\nu) = \int f(x)\varphi(x-b)e^{-i\nu x}dx = \langle f(x), O_{b,\nu}\rangle \tag{2.24}$$

where $O_{b,\nu} = \overline{\varphi(x-b)}e^{-i\nu x}$. $\Phi_b(f)(\nu)$ is called the Short Time Fourier Transform (STFT) of $f(x)$.

The resulting expansion is a function of two parameters: frequency and time shift. Equation (2.24) appears to be very ideal since (roughly speaking), there are two element in the equation, location or time shift(b) and frequency (ν). The main property is that the window size is fixed with respect to frequency which produces a rectangular partitioning of the time-frequency plane, as shown in figure 2.7

In order to grasp how stable equation (2.24) is, one needs to know the concept of time and frequency *localisation*, which is what next subsection tries to do in the simplest possible form.

2.3.1 Time and Frequency Localization

Recall the definition of $f(x)$ with finite energy from equation (2.4):

[1]Named after Dennis Gabor, the famous physicist who won the Nobel Prize in Physics in 1971.

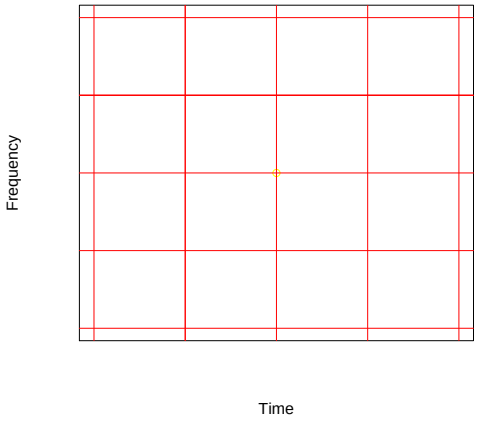

Figure 2.7: The balanced resolution between time and frequency by using the Short Time Fourier transform.

$$\int_{-\infty}^{\infty} |f(x)|^2 dx < \infty \tag{2.25}$$

If $f(x)$ has finite energy then it is well localised in the time domain if it satisfies some decay properties away from a fixed value. For example, $f(x)$ is polynomially localised near $x = x_0$ if (examples are all from (Carmona et al., 1998)):

$$|f(x)| < K \frac{1}{(1 + (x - x_0)^2)^{k/2}}, x \in \mathbb{R} \tag{2.26}$$

for some positive constants K and k. Or $f(x)$ is exponentially localised if:

$$|f(x)| < K e^{-\alpha|x-x_0|}, x \in \mathbb{R} \tag{2.27}$$

for some positive constants K and α. In equations (2.26) and (2.27) x can be replaced by ν to be regarded as exponential and polynomial frequency localisations.

2.3.2 Heisenberg's Inequality Principle

Now that the concepts of time and frequency localisations are defined, the stability of equation (2.24) can be explained through the bridge between time and frequency

localisation which is what Heisenberg's Inequality Principle is about. According to Heisenberg (Heisenberg, 1930), it is not possible to achieve *optimal* localisation simultaneously in time and in the frequency domain (look at equation (2.24)). Improving a function's time localisation will end up weakening its frequency localisation. This fact is a consequence of the well known *Heisenberg Uncertainty Principle*.

Consider $f \in L^2(\mathbb{R})$, and assume that its derivatives f' and its Fourier transform are also in $L^2(\mathbb{R})$. If the time and frequency averages are defined as:

$$\overline{x} = \frac{1}{||f||^2} \int x|f(x)|^2 dx, \tag{2.28}$$

and

$$\overline{\nu} = \frac{1}{||\hat{f}||^2} \int \nu|\hat{f}(\nu)|^2 d\nu, \tag{2.29}$$

where the corresponding time and frequency variances are:

$$\Delta_x = \frac{1}{||f||} \sqrt{\int (x - (\overline{x}))^2 |f(x)|^2 dx} \tag{2.30}$$

and

$$\Delta_\nu = \frac{1}{||(\hat{f})||} \sqrt{\int (\nu - (\overline{\nu}))^2 |\hat{f}(\nu)|^2 d\nu} \tag{2.31}$$

The Heisenberg uncertainty principle states that:

$$\Delta_x \Delta_\nu \geq \frac{1}{2} \tag{2.32}$$

Equation 2.32 shows that an improvement of the time localisation (a decrease of Δ_x) will result in a deterioration in the frequency localisation (an increase of Δ_ν). [1]. It is crucial to point out that achieving an infinitely precise description of the time-frequency content of a time series/signal is *impossible* because of the Heisenberg uncertainty principle [(Carmona et al., 1998), (Reed and Simon, 1976)] and there is always a compromise between time localisation and frequency localisation. Mathematicians, engineers, and physicists have all been concerned with finding an ideal (suited for their application)

[1] The proof of equation 2.32 is not central to the discussion of this thesis and is not provided here, a complete proof can be found in (Reed and Simon, 1976).

time-frequency representation of signals/time series in terms of a time variable and a frequency (or scale) variable simultaneously.

It should be now obvious that time-frequency representations are not unique. There are different ways of describing the information content of a signal in time and frequency domain. Depending on the application, this can change.

2.3.3 Shortcomings of STFT

Equation (2.24) which is the windowed Fourier transform means that at every point in the time-frequency domain, a window is chosen which is in accordance with a specific time and frequency. The duration and bandwidth of the window and thus the resolution do not change. Hence, the accuracy of STFT is limited by the size and the shape of the window. If one uses many time intervals will end up with good time resolution but the very short time of each window would not give good frequency resolution, hence loosing frequency information. The resolution in the time and frequency domains depends only on the form of the window and, by the Heisenberg uncertainty principle, it is impossible to achieve an optimal resolution simultaneously in the time and frequency domains. Also another drawback of the STFT is that it is applying the Fourier transform to bits of the observed time series and it will not be able to capture the events that fall outside the width of the window.

For a more extensive and detailed studies on the drawbacks of Short Time Fourier Transform please see: (Carmona et al., 1998), (Ramsey, 2002), and (Gao and Yan, 2011).

2.4 Wavelet Analysis

The wavelet transform is a powerful mathematical tool that is receiving more and more attention in the statistical and financial communities. The power of wavelets is their ability to analyse (decompose) features which vary over both time and scale. Wavelet transform instead of partitioning the time-frequency plane into a square grid like STFT, adaptively partitions the time-frequency plane so that it better captures the range of low-frequency to high-frequency events. Figure 2.8 shows how the time-frequency plane is partitioned (in comparison to figures (2.6), (2.5), and (2.7)).

To overcome the fixed time-frequency partitioning in STFT, wavelet transform exploits a basic function by (often in the literature called *analysing* or *mother* wavelet), shifting (translating) and scaling (dilating) so that it captures all the features local in time and frequency. Wavelet transform does not violate the Heisenberg's principle, it rather gets round it intelligently. The time-frequency partition analyses high-frequency events with good time resolution and low-frequency features with good frequency resolution which, is in contrast to figure (2.7) where all frequencies were analysed with the same fixed-width window. As frequency increases, the number of partitions in the time domain also increases. This enables a balanced resolution in time and frequency simultaneously. Recall that in STFT, if the focus is on high frequency characteristics then the time resolution is lost and vice versa. However, in wavelet transform one is enabled to have both. This is how wavelet transform overcomes the limitations in STFT and also does not contradict Heisenberg's principle.

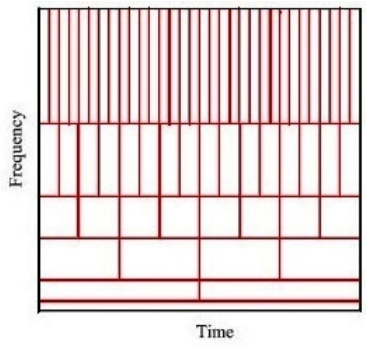

Figure 2.8: The partitioning of the time-frequency plane according to a wavelet transform.

To be precise, Fourier transform and STFT deal with frequency, however, wavelet transform deals with scaling. Roughly speaking, scale is inversely proportional to the frequency interval.

If the scale parameter in wavelet transform increases, then it increases the time support, drops off in frequency domain, hence capturing lower frequencies (figure (2.9, a)). On the other hand if the scale parameter decreases, then the wavelet basis decreases the time support and covers high frequencies (figure (2.9, d)). Figure (2.9) shows the Haar wavelet function, shifted towards time axis and scaled towards the frequency axis.

For an introduction to wavelet analysis that does not go through the rigorous maths

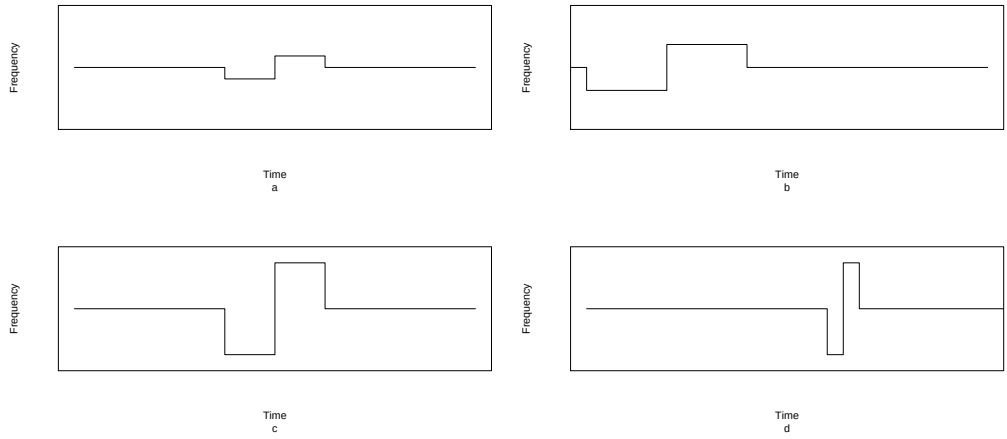

Figure 2.9: Haar Wavelets, shifted and dilated across time and scale.

(Hubbard, 1998) gives a non-technical introduction to wavelets. For good sources with moderate levels of technicality (Ramsey, 1999), (Schleicher, 2002), (Crowley, 2007), and (Gallegati, 2008) are some easy-to-follow sources to start with. For a more extensive and thorough introduction to the theory of wavelets and wavelet transforms, see (Chui, 1992), (Daubechies, 1992), (Carmona et al., 1998) (Mallat, 1998) and (Percival and Walden, 2000) (specifically for time series analysis) which are mostly used as main references.

Recently, there have been a growing number of applications of wavelets in areas from theoretical econometrics to Capital Asset Pricing Model. (Zhang and Benveniste, 1992) applies wavelets to Artificial Neural Networks (ANN), (Nason, 2008) discusses the applications of wavelet in statistics using **R**. (Ramsey and Lampart, 1998) analyse the scale by scale relationship between expenditure and income for upper middle income countries, while (Gencay et al., 2002) apply wavelets on a much wider scale and multiple empirical applications in finance. (Vuorenmaa, 2004) uses wavelets for a volatility analysis of 5-minute Helsinki Stock Exchange, and (Fan and Wang, 2007) apply wavelets for removing the noise from high frequency data when calculating realised variance. While (Power and Turvey, 2010) apply wavelets for analysing the volatility of energy markets, (Vacha and Barunik, 2012) employs wavelets for studying the co-movement of energy commodities. Overall, the use of wavelets have started to appear more frequently in empirical analysis of high frequency quadratic variations.

2.4.1 Wavelet Functions

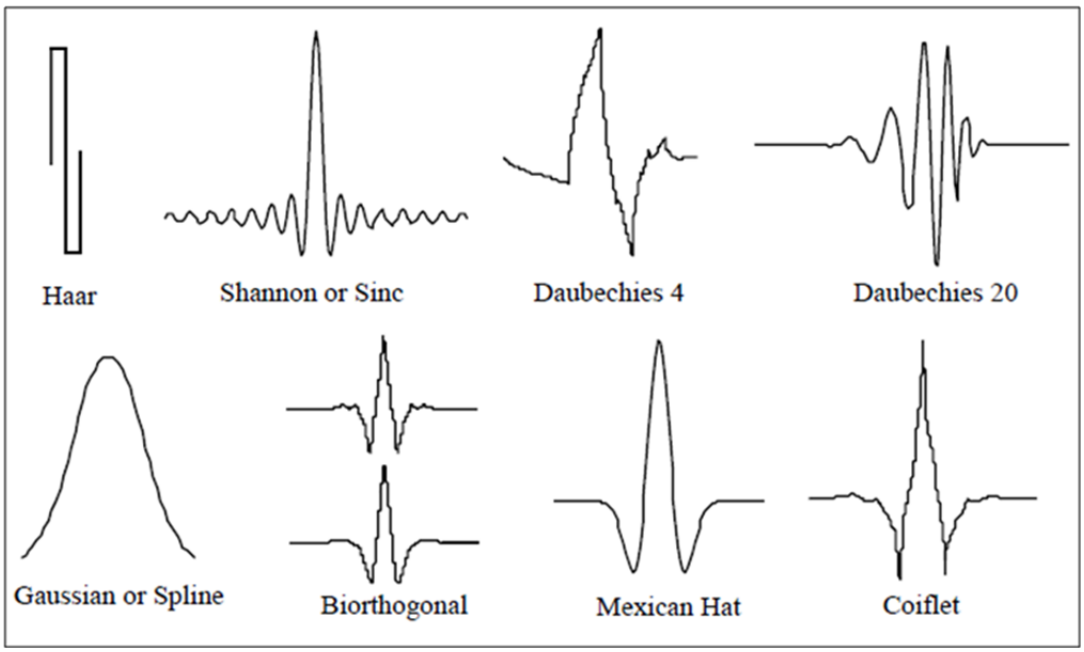

Figure 2.10: Examples of different types of wavelets. Pictures are drawn in Matlab, Wavelet Toolbox.

A *wavelet* as the name suggests, is a *small wave*. A small wave can decay and grow in a limited time period which is in contrast with sinwave which keeps on oscillates up and down.

A wavelet $\psi(t)$ is simply a function of time t if it satisfies two basic rules (the integral of ψ is zero):

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \tag{2.33}$$

A secondary condition imposed on a wavelet condition is unit energy[1]; that is

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1 \tag{2.34}$$

Just for the sake of comparison, a sinwave is not an appropriate candidate as

[1] The energy of a function is defined to be the squared function integrated over its domain. This condition should hold so that the inverse of wavelet transforms exists.

$\int_{-\infty}^{\infty} |\sin(t)|^2 dt$, diverges.

Haar wavelet function is the oldest, the most famous, and the simplest wavelet basis, named after Alfred Haar:

$$\psi^{Haar}(t) = \begin{cases} 1 & on[0,1/2] \\ -1 & on[1/2,1] \\ 0 & Otherwise \end{cases}$$

It is straightforward to check whether equations (2.33) and (2.34) hold.

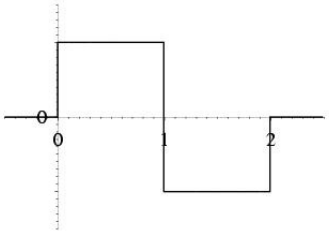

Figure 2.11: Haar Basis (mother) Function.

$$s = 2^{-j} and u = k2^{-j} \tag{2.35}$$

Equations (2.33) and (2.34) are two very basic rules of wavelets and guarantee that wavelet function will have nonzero entries, but all departures from zero must cancel out. This means that the function ψ has to wiggle up and down the time axis, i.e. it must behave like a wave; this, together with the assumed decaying property, justifies the choice of the term wavelet (in French, ondelette). For wavelets to be of practical use, they should obey a basic rule, known as the wavelet admissibility condition. A wavelet ψ is admissible if its Fourier transform satisfies:

$$\psi(\nu) = \int_{-\infty}^{\infty} {}^{-i2\pi\nu t} dt \tag{2.36}$$

This condition ensures that $\psi(\nu)$ (where $\hat{\psi}(\nu)$ is the Fourier transform, of $\psi(t)$) goes to zero quickly as $\nu \to 0$ $Mallat$ (1998)), that is:

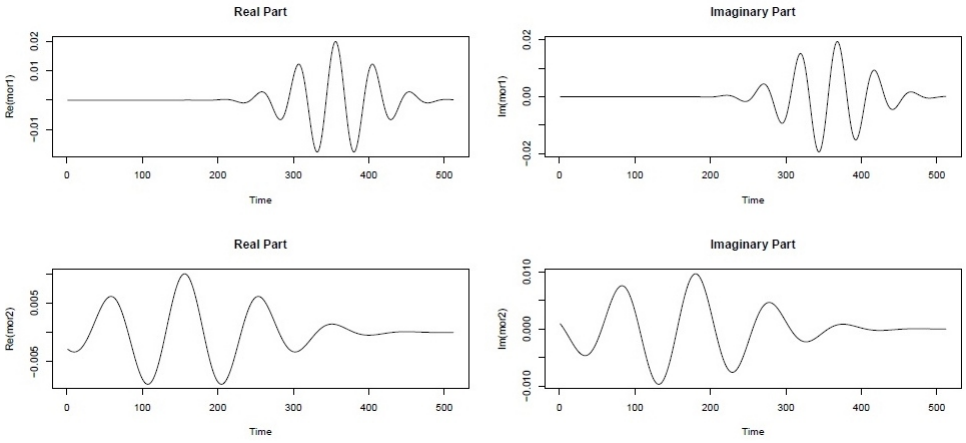

Figure 2.12: Two Morlet wavelets. The first one at given scale 50, and location 356. The second one at given scale 100, and location 100. Please see the appendix for the codes.

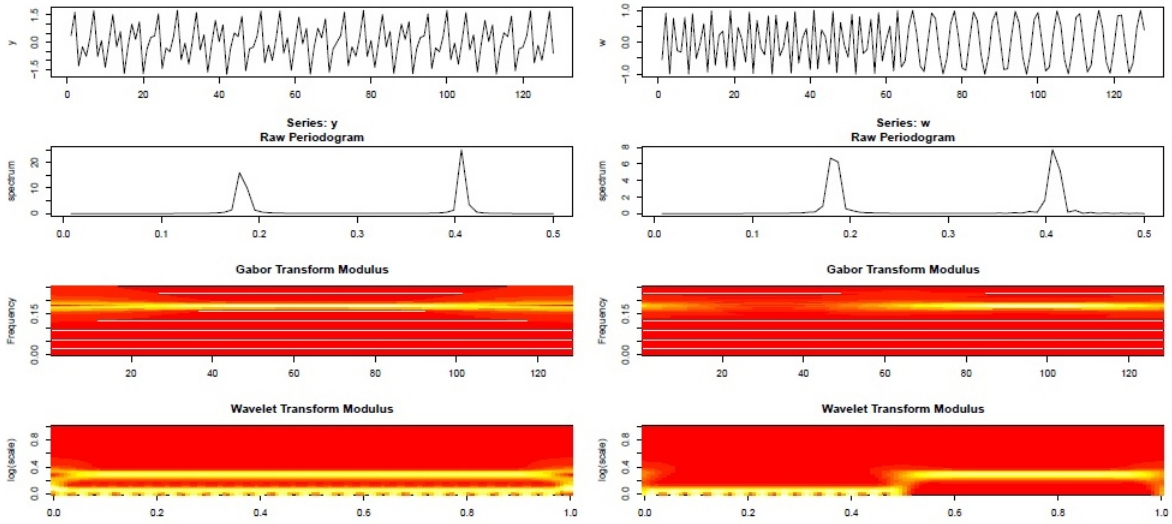

Figure 2.13: A complete version of figure (2.4). Plots of two signals against time, Spectrum against frequency, Gabor Transform (STFT) and CWT transforms. Signal (top, left) consists of superposition of two frequencies $sin(10t)$ and $sin(20t)$, and the second (top, right) consists of the same two frequencies each applied separately over half of the signal duration. Codes are given in appendix.

$$0 < C_\psi = \int_0^\infty \frac{\psi(f)}{f} df < \infty \tag{2.37}$$

Daubechies wavelets (another type of wavelets commonly used) which are a family of compactly supported wavelets satisfy an additional condition which is based on vanishing moments, that is:

$$\int t^m \psi(t) dt = 0 \tag{2.38}$$

where where $m = 0, 1, ..., M-1$, (Daubechies, 1992). Another classical example of wavelet family is Morlet wavelet, a continuous-time wavelet:

$$\psi^{Morlet}(t) = \frac{1}{2\pi} e^{-i\nu_0 t} e - \frac{t^2}{2} \tag{2.39}$$

where $i = \sqrt{-1}$ and is an imaginary number and ν_0 is the central frequency of the wavelet. Figure (2.12) shows two Morlet wavelets, with different ν. Strictly speaking, the Morlet wavelet is not an admissible wavelet, since it is not of integral zero. However, for ν large enough, the integral of ψ is small enough to ensure that for all practical purposes, it could be used numerically as if it were a wavelet.

Some of other examples of continuous-time wavelet functions (some which are shown in figure (2.10)) are related to derivatives of Gaussian probability density function. The normalised wavelet which is the first derivative of Gaussian PDF, and also Mexican hat wavelet which is the second derivative of Gaussian PDF.

2.4.2 Continuous Transform or Discrete Transform

The continuous wavelet transform (CWT) is a function of translation (shift) and scale (dilation) $W_{u,s}$ which is obtained by developing the function onto a particular wavelet via:

$$W_{u,s} = \prec x, \psi_{u,s} \succ = \int_{-\infty}^{+\infty} x(t) \psi_{u,s}(t) dt \tag{2.40}$$

where

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi(\frac{t-u}{s}), s \in \mathbb{R}^+, u \in \mathbb{R}, s \neq 0 \tag{2.41}$$

is the shifted by u (the position of the wavelet in the time domain), and scaled by s (the position of the wavelet in the frequency domain) compared to the original wavelet function. Compare figure (2.9) (which shows dilations and translation of the original Haar basis function), with figure (2.11) (which shows the Haar basis function). The translation of a wavelet function $\psi(t-u)$ simply shifts its u range units to the right(depending on the sign of u), while a dilation of the function $\psi(\frac{t}{s})$ expands its range by a multiplicative factor s (if $|s| > 1$ stretching and if $|s| < 1$ compressing, remember that s has to be positive). The concepts of translation and dilations might seem unfamiliar to the world of conventional econometrics. In the literature u is a factor of translation and s is the factor for dilation. Translation means shifting and dilation means scaling and it is just a matter of using different terminologies.

The resulting wavelet coefficients ($\psi_{u,s}(t)$) are a function of these two parameters, location (u) and scale(s)[1]. The original function is a function of one parameter (t here) but the wavelet coefficients are a function of two parameters s and u. In equation (2.41), the term $\frac{1}{\sqrt{s}}$ guarantees that equation (2.34) holds (in other words it makes sure that the transform is invertible).

Applying shifted and scaled wavelets to a function, breaks down the complicated structure of the function into simpler components. This is what is called analysing or decomposing a function into sub-functions of simpler dynamics. If a wavelet satisfies the admissibility condition (equation (2.37)), then an inverse wavelet transform can re-produce the original function (same as it was in Fourier Transform):

$$x(t) = \frac{1}{c_\psi}\int_{s=0}^{s=+\infty}\int_{u=-\infty}^{u=+\infty}\psi_{s,u}(t)W(u,s)duds\frac{1}{s^2} \tag{2.42}$$

Continuous Wavelet Transform (CWT) is a function of two parameters and therefore maps the original series into a function of u and s. CWT is often use in image processing/denoising where the original information of the object is observed in two dimensions, time and space. For the sake of comparison, recall figure (2.4), which shows two time series: the first one consisting of two frequencies ($sin10t$ and $sin20t$)

[1] In the literature some times ψ is called the "mother wavelet" and a family of $\psi_{u,s}$ are called "wavelet daughters"! (Hubbard, 1998)

for the entire duration of the signal ($t = 60$), and the second one consisting of the same frequencies ($sin10t$ and $sin20t$), but each is applied separately for half of the signal duration: for the first half ($sin10t$) and for the second half ($sin20t$). The spectrum was quite incapable of realising the difference between the two signals. Figure 2.13, re-plots the figures adding, Short Time Fourier Transform (Gabor) and Continuous Wavelet Transform (CWT). It is obvious that even though these two signals are fairly simple, Gabor transform still is not capable to verify where in time the second signal changes its frequency. On the other hand, CWT clearly is capable of realising the differences. Morlet wavelet has been used to produce the CWT.

CWT is obtained by continuously shifting a continuously scalable function and then calculates the correlation between the the signal and the shifted, scaled functions. Therefore lots of correlations are built in. In economics and finance all objects are observed in one dimension : time, and thus CWT (if applies on one dimensional signals) gives information simultaneously on time and frequency which result in containing a high amount of extra (redundant) information. CWT transfers a one-dimensional signal to a two-dimensional time-scale representation and at all points in the time frequency plane, the wavelet coefficients contain more information than necessary for the perfect reconstruction property to hold. A smaller number of scales with a varying number of wavelet coefficients at each scale can be used, instead of using continuous parameters u and s in order to avoid redundancy. For example, in figure (2.13), the change in frequency is observed to some degree at every scale, enough information would have been capable to detect the change in the dynamics of the signal. And that is how the concept of Discrete Wavelet Transform (DWT) enters the picture. one can think of DWT as a discretization of the CWT throughout sampling specific wavelet coefficients. A critical sampling of the CWT is obtained via (Vidakovic, 1999), and (Percival and Walden, 2000):

where j and k are integers representing the set of discrete shifts and scales. The term "critical sampling" is used to represent the minimum number of coefficients sampled from CWT to ensure that all information present in the original function is retained by the wavelet coefficients. If $x(t)$ is a signal with finite energy:

$$\int x^2(t)dt < \infty \tag{2.43}$$

then can be analysed by projecting it with the wavelet coefficients. The wavelet coefficient is equal to:

$$w_{u,s} = \prec x, \psi_{u,s} \succ = \int_{-\infty}^{+\infty} x(t)\psi_{u,s}(t)dt = 2^{\frac{j}{2}} \int_{-\infty}^{+\infty} x(t)\psi(2^j t - k)dt \tag{2.44}$$

The wavelet coefficient $w_{u,s}$ is both localized in time (location) and scale, that is: each basis function depends on two parameters, the scale s and locations u, whereas the Fourier basis functions only depend on a single parameter f frequency which is $\frac{1}{s}$ (recall that it was mentioned scale is inversely proportional to frequency of an interval). The wavelet coefficient $w_{u,s}$ in equation (2.44) is sometimes written as $w_{k,j}$. However, in this thesis $w_{u,s}$ means (Discrete Wavelet Transforms) DWT unless otherwise stated.

Without going into to extensive mathematics, it is worth mentioning that, if the frequency-domain support of $w_{s,u}$ at time u is $[-2^{-s}, -2^{-s-1}) \cup (2^{-s-1}, s^{-s}]$, then the time-domain support of $w_{s,u}$ is $[u2^{-s}, (u+1)2^{-s}]$. Thus the wavelet function has the ability to zoom in and out according to scale (s) (please see (Daubechies, 1992), (Mallat, 1998), (Percival and Walden, 2000) and (Nason, 2008)).

Signal $x(t)$ can be expanded via orthogonal wavelet basis:

$$x(t) = \sum_{s=0}^{s=+\infty} \sum_{u=-\infty}^{u=+\infty} w_{s,u}\psi_{s,u}(t) \tag{2.45}$$

And equation (2.45) can be interpreted as an aggregation of details across all scales which will be used in chapter four.

2.4.3 The Discrete Wavelet Coefficients

In most of cases, there are no functions to represent the wavelets details and the scaling functions in the time domain. Therefore, iterative procedures are used for generate them. Let $h_l = \{h_0, ..., h_{L-1}\}$ denote the wavelet basis (high-pass) filter and let $g_l = \{g_0, ..., g_{L-1}\}$ be the scaling (low-pass) filter, defined via[1]:

$$g_m = (-1)^{m+1} h_{L-1-m} \tag{2.46}$$

[1] A high-pass filter passes high-frequency dynamics but reduces the amplitude of low dynamics. Low pass filter passes low-frequency dynamics but reduces the amplitude of high frequency dynamics

This relationship which is often known as a quadrature mirror filter is a filter used for splitting an input signal into two bands. The resulting high-pass and low-pass signals are often reduced by a factor of 2, giving a critically sampled two-channel representation of the original signal. Orthogonal wavelets satisfy a quadrature mirror filter relationship. In the literature mostly the low-pass filter coefficients are obtained through equation 2.46 (Daubechies, 1992) and (Mallat, 1998).

The condition in equation 2.34 should also be satisfied for the wavelet filters $\{h_l\}$: $\sum_{l=0}^{l=L-1} h_l^2 = 1$ (has unit energy). Also following equation 2.33 ($\int \psi(t)dt = 0$)), $\sum_{l=0}^{l=L-1} h_l = 0$ (it integrates to zero) and $\sum_{l=0}^{l=L-1} h_l h_{l+2n} = 0$ (is orthogonal to its even shifts).

If $x(t)$ is a signal with a dyadic length ($N = 2^J$) of observations. The vector of discrete wavelet coefficients ω is obtained via

$$\omega = \mathbf{W}x \tag{2.47}$$

Where $\mathbf{W}$ is an $N \times N$ orthonormal matrix defining the DWT. The matrix $\mathbf{W}$ is formed of the wavelet and scaling filter coefficients. Equation 2.47 becomes more clear if written as:

$$\mathbf{w} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_J \\ \mathbf{v}_J \end{bmatrix} \tag{2.48}$$

and

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \vdots \\ \mathbf{W}_J \\ \mathbf{V}_J \end{bmatrix} \tag{2.49}$$

where w_j is a column vector of wavelet coefficients with length $\frac{N}{2^j}$ which is associated with changes on a scale of length $s_j = 2^{j-1}$ and v_j is a column vector of scaling

coefficient which is associated with averages on a scale of length $s_j = 2^{J-1}$, hence W_j and V_j are sub-matrices of size $(\frac{N}{2^j} \times N)$. To further simplify equation 2.49, let:

$$\mathbf{h}_1 = [h_{1,N-1}, h_{1,N-2}, ..., h_{1,1}, h_{1,0}]^T \tag{2.50}$$

be the vector of zero padded unit scale wavelet filter coefficients in reverse order. That is, the coefficients $h_{1,0}, ..., h_{1,L-1}$ are taken from an appropriate orthonormal wavelet family of length L, and all values that $L < t < N$ are defined to be zero. Now if shift h_1 shifts by factors of two:

$$\mathbf{h}_1^{(2)} = [h_{1,1}, h_{1,0}, h_{1,N-1}, h_{1,N-2}, ..., h_{1,3}, h_{1,2}]^T$$
$$\mathbf{h}_1^{(4)} = [h_{1,3}, h_{1,2}, h_{1,1}, h_{1,0}, h_{1,N-1}, h_{1,N-2}, ..., h_{1,5}, h_{1,4}]^T$$

and so on. And then W_1 in equation 2.49 is the collection of $N/2$ shifted versions of $\mathbf{h}_1$:

$$\mathbf{W}_1 = [h_1^{(2)}, h_1^{(4)}, ..., h_1^{(N/2-1)}, h_1]^T \tag{2.51}$$

Let $\mathbf{h}_2$ be the vector of zero padded scale 2 wavelet filter coefficients defined similarly to the process above, then sub-matrix $\mathbf{W}_2$ is constructed by shifting the vector $\mathbf{h}_2$ by factors of 4. This way of construction is repeated to form the matrices $\mathbf{W}_j$ by shifting the vector h_j by factors of 2^j. The matrix $\mathbf{V}_J$ is simply a column vector whose elements are all equal to $1/\sqrt{N}$, (see (McCoy and Walden, 1996) and (Percival and Walden, 2000)).

Finally the $N \times N$ dimensional matrix $\mathbf{W}$ is structured by the sub-matrices $\mathbf{W}_1, \mathbf{W}_2, ..., \mathbf{W}_J$ and $\mathbf{V}_J$ in equation 2.49. A vector of wavelet coefficients is associated with changes at a particular scale. Level j wavelet coefficients $w_{j,t}$ are associated with the scale $= 2^j$. This means that each wavelet coefficient is constructed using a difference of two weighted averages, each one of length $_j$. Applying the DWT to a stochastic process produces a decomposition on a scale by scale basis which is the unique characteristic of wavelets and becomes very useful in chapter four.

Haar wavelet is nearly identical to the familiar first difference operator. Applying the Haar DWT to the random walk x_t will produce stationary wavelet coefficients at all levels. If x_t is formed by a double sum of a Gaussian white noise process, then

$$x_t = \sum_{u=1}^{t} \sum_{v=1}^{u} \epsilon_v, t \geq 1, \tag{2.52}$$

The first difference of x_t is a random walk and the second difference is a Gaussian noise. $D(4)$ wavelet filter is closely related to the second difference operator. In this thesis apart from DWT, another type of discrete wavelets are used, known as Maximal Overlap Discrete Wavelet Transform (MODWT) which has a different sampling algorithm and is more redundant compared to DWT.

2.4.4 The Maximal Overlap Discrete Wavelet Transform

An alternative wavelet transform to DWT is the maximal overlap discrete wavelet transform (MODWT)(the term *maximal overlap* comes from the relationship of the MODWT with estimators of the Allan variance (Percival and Guttorp, 1994)). MODWT is seen in the literature by different names, such as the *Stationary DWT* in (Nason, 2008) or *Translation Invariant DWT* in (Coifman and Donoho, 1995). MODWT gives up orthogonality in DWT but gains features that DWT does not have. Briefly the important differences between these DWT and MODWT can be summarised as (Percival and Guttorp, 1994):

- MODWT can handle signals of any length, while DWT is restricted to signals of dyadic length. Although often in DWT, if the length of the original series is not far from a dyadic number, *0*s are added to the signal.
- MODWT is invariant to circularly shifting the original time series. When the signal is shifted by an integer, the MODWT basis and scaling coefficients will be shifted by the same amount. This property does not hold for DWT.
- MODWT variance estimator is asymptotically more efficient than the same DWT variance estimator (more on this here: (Percival and Walden, 2000)).
- MODWT multiresolution analysis can be aligned with the features in the time series (Which will be used in chapter four) (Percival and Walden, 2000).

If $x(t)$ is a signal of any length N (not necessarily dyadic), the length $(J+1)N$ vector of MODWT coefficients ($\widetilde{W}$) is obtained via:

$$\widetilde{\omega} = \widetilde{W}x \tag{2.53}$$

where $\widetilde{W}$ is a $(J+1)N \times N$ matrix and similar to DWT, the matrix $\widetilde{W}$ can be made up of $J+1$ submatrices, each of them $N \times N$:

$$\widetilde{\mathbf{w}} = \begin{bmatrix} \widetilde{\mathbf{w}}_1 \\ \widetilde{\mathbf{w}}_2 \\ \vdots \\ \widetilde{\mathbf{w}}_J \\ \widetilde{\mathbf{v}}_J \end{bmatrix} \tag{2.54}$$

and

$$\widetilde{\mathbf{W}} = \begin{bmatrix} \widetilde{\mathbf{W}}_1 \\ \widetilde{\mathbf{W}}_2 \\ \vdots \\ \widetilde{\mathbf{W}}_J \\ \widetilde{\mathbf{V}}_J \end{bmatrix} \tag{2.55}$$

where $\widetilde{\omega}_j$ is a vector of wavelet coefficients of length $N/2^{j-1}$ associated with changes on a scale of length $s = 2^{j-1}$ and $\widetilde{v}_J$ is a length $N/2^J$ vector of scaling coefficients with averages on a scale of length $s = 2^{J-1}$. MODWT uses the rescaled filters and does not use wavelet and scaling filters like DWT does [(Gencay et al., 2002), (Mallat, 1998), and (Percival and Walden, 2000) have slightly different approach towards MODWT scaling and detail filters].

$$\widetilde{h}_j = h_j/2^j \tag{2.56}$$

and

$$\widetilde{g}_J = g_J/2^J \tag{2.57}$$

The $N \times N$ dimensional submatrix $\widetilde{W}_1$ is constructed by shifting the rescaled wavelet filter by integer units to the right:

$$\widetilde{W}_1 = [\widetilde{h}_1^{(1)}, \widetilde{h}_1^{(2)}, ..., \widetilde{h}_1^{(N-1)}, \widetilde{h}_1]^T \tag{2.58}$$

The remaining submatrices $\widetilde{W}_2, ..., \widetilde{W}_J$ are formed similarly to the ones in DWT except that $\widetilde{h}_1$ is replaced by $\widetilde{h}_j$.

2.4.5 Multiresolution Analysis

A multiresolution of $L^2(\mathbb{R})$ is defined as a sequence of subspaces (closed subspaces) V_j of $L^2(\mathbb{R})$ [(Daubechies, 1992), (Mallat, 1998), and (Pollock, 2012)] with the following properties:

- $V_j \supset V_{j+1}$,
- $v(x) \in V_j \iff v(2x) \in V_{j+1}$,
- $v(x) \in V_0 \iff v(x+1) \in V_0$,
- $\bigcup_{j=-\infty}^{+\infty} V_j$ is dense in $L^2(\mathbb{R})$ and $\bigcap_{j=-\infty}^{+\infty} V_j = 0$,

That is:

$$V_0 \supset V_1 \supset ... \supset V_n \tag{2.59}$$

The jth stage of the process, which generates V_j , also generates the accompanying space W_j of wavelet functions, which is its orthogonal complement within V_{j-1}. The complete process can be summarised by displaying the successive decompositions of V_0:

$$\begin{aligned} V_0 &= W_1 \oplus V_1 \\ &= W_1 \oplus W_2 \oplus V_2 \\ &\quad ... \\ &= W_1 \oplus W_2 \oplus ... \oplus W_n \oplus V_n \end{aligned} \tag{2.61}$$

Using MODWT (or DWT), an additive decomposition of a series of observations can be formed:

$$x_t = \sum_{j=1}^{J+1} \widetilde{d}_{j,t}, \quad t = 0, ..., N-1 \tag{2.62}$$

where $\widetilde{d}_{j,t}$ is the $t-th$ element of $\widetilde{d}_j = \widetilde{\mathbf{W}}_j^T \widetilde{w}_j$ for $j = 0, ..., N-1$. That is, each signal is a linear combination of wavelet detail coefficients. Let $\widetilde{s}_{J,t} = \sum_{k=j+1}^{J+1} \widetilde{d}_{k,t}$ define the $j-th$ level wavelet smooth for $0 \leq j \leq J$, where $\widetilde{s}_{J+1,t}$ is defined to be a vector of zeros. While wavelet detail $\widetilde{d}_{k,t}$ is associated with variations at a particular scale, $\widetilde{s}_{J+1,t}$ is a cumulative sum of these variations. Wavelet detail $\widetilde{r}_{j,t} = \sum_{k=1}^{j+1} \widetilde{d}_{k,t}$ shows the lower scale details (high frequency features). A signal can be decomposed through a wavelet smooth and rough via:

$$x_t = \widetilde{s}_{J,t} + \widetilde{r}_{j,t} \tag{2.63}$$

Multiresolution analysis (MRA) is used for filtering out the seasonalities and it is used in chapter four to decompose the series into different sub-signals each showing the original signal at different scales. The use of MRA has proved to be very popular in economics and finance, since it breaks the complication of observed information into simpler building blocks [to name a few: Gencay et al. (2001b), Gencay et al. (2003), Kim and In (2010)]

2.4.6 The Wavelet Variance

The orthonormality of the matrix $\mathbf{W}$ (in equation 2.47) implies that the DWT is an energy (variance) preserving transform (recall that, the term "energy" means the sum of squared coefficients of a vector). That is, the variance of the original time series is perfectly captured by the variance of the coefficients from the Discrete Wavelet Transform:

$$||\mathbf{w}||^2 = \sum_{j=1}^{J} \sum_{t=0}^{N/2^j-1} w_{j,t}^2 + v_{J,0}^2 = \sum_{t=0}^{N-1} x_t^2 = ||\mathbf{x}||^2$$

which means that $||x||^2$ can be decomposed on a scale by scale basis via:

$$||\mathbf{x}||^2 = \sum_{j=1}^{J} \|\mathbf{w}_j\|^2 + \|\mathbf{v}_J\|^2$$

where $\|w_j\|^2$ is the energy (proportional to variance) of x due to changes at scale s_j and $\|v_J\|^2$ is the information due to changes at scale s_J and higher.

(Percival and Walden, 2000) proved that the MODWT is an energy preserving transform too. However while a variance decomposition using the wavelet details and smooth is valid for DWT but it is not the case for MODWT, but MODWT can still be used providing MODWT coefficients are used and not the wavelet details. Let $\widetilde{w}$ be the MODWT coefficient vectors, respectively of series x (a mean zero series), we can then decompose the variance of the original time series using MODWT and have the following relationship:

$$\|\mathbf{x}\|^2 = \sum_{j=1}^{J} \|\widetilde{\mathbf{w}_j}\|^2 + \|\widetilde{\mathbf{v}}_J\|^2 \tag{2.64}$$

Equation (2.64) provides a decomposition of variance of the original series and MODWT wavelet coefficients and will be used in chapter four.

Applying the DWT on to a stochastic process produces a decomposition of the process on a scale-by-scale basis. Following (Percival and Walden, 2000), suppose that y_t is a stochastic process that is not necessarily a stationary process. The time-varying wavelet variance for for y_t is defined as:

$$\sigma^2_{y,t}(S_j) = \frac{1}{2S_j} Var(w_{j,t}) \tag{2.65}$$

which is the variance of the scale S_j wavelet coefficient $w_{j,t}$. If the stochastic process is covariance stationary , then:

$$\sigma^2_{y}(S_j) = \frac{1}{2S_j} Var(w_{j,t}) \tag{2.66}$$

For the sake of comparison, recall that Fourier transform decomposes the variance of a process on a frequency-by-frequency basis. Wavelet variance decomposes the variance of the process on a scale-by-scale basis:

$$\sum_{j=1}^{\infty} \sigma^2_{y}(S_j) = Var(y_t) \tag{2.67}$$

The scale S_j is associated with the frequency interval $[\frac{1}{2^{j+1}}, \frac{1}{2^j}]$, using this property an approximate relation between the wavelet variance and the spectral density function:

$$\sigma_y^2(S_j) \approx 2 \int_{\frac{1}{2^{j+1}}}^{\frac{1}{2^j}} S_y(\nu) d\nu \tag{2.68}$$

In order to estimate the wavelet variance, (using (Percival and Walden, 2000)'s approach) suppose y_t which is a stochastic process of any arbitrary length N and the partial MODWT of order $J_P < log_2(N)$ to produce the length $(J_p + 1)N$ vector of wavelet coefficients $\widetilde{w}$. An unbiased estimator of the wavelet variance is based on the MODWT using:

$$\sigma_y^2(S_j) = \frac{1}{\widetilde{N}_j} \sum_{t=L_j-1}^{N-1} 2\widetilde{w_{j,t}^2} \tag{2.69}$$

where $L_j = (2^j - 1)(L - 1) + 1$ is the length of the scale S_j wavelet filter and $\widetilde{N}_j = N - L_j + 1$ is the number of coefficients unaffected by the boundary. (Gencay et al., 2002) and (Percival and Walden, 2000) show that the estimator $\widetilde{\sigma}_y^2(S_j)$ is asymptotically Gaussian distributed with mean $\sigma_y^2(S_j)$ and variance $S_{\overline{w},t} = w_{j,t}^2$ evaluated at frequency zero and is given by:

$$S_{\overline{w},0}^G = 2 \int_{-1/2}^{1/2} S_j^2(\nu) d\nu \tag{2.70}$$

(Percival and Walden, 2000) and (Gencay et al., 2002) use equation (2.70) to compute confidence intervals under both Gaussian and non-Gaussian assumption for y_t which is not mentioned here.

3

Volatility Modeling: Using Daily Observations

3.1 Introduction

In the history of humankind, financial markets are to be known as one of the most exclusive and complicated systems. Financial markets are combinations of interactions of thousands of individuals in different geographical locations who have different time horizons, risk attitudes, currencies, information access, and different utility functions. Each of these factors map heterogeneities of their own natures onto financial markets.

The transactions happen at different time scales. On the one hand traders have access to ultra-high frequency data with time stamps at nanosecond resolutions in which placing orders are controlled by computer algorithms. On the other hand, day traders or trading desks of banks calculate the profit and loss daily and their positions are fairly liquid, it therefore makes sense to calculate the risk over a time horizon of one trading day. For an investment portfolio held by a pension fund, a time horizon of one month may be chosen as the portfolio is traded less actively, some of the assets in the portfolio are less liquid and also the performance of pension fund portfolios is often monitored monthly. An appropriate choice for the time horizon depends on the application and there is no "golden" horizon that works for everyone.

To add to the complication, the valuation of financial assets are driven from news as diverse as short-run weather forecasts (Roll, 1984) which are only relevant for a few

hours, and macroeconomic technological revolutions that may take years to come in to effect (Pastor and Veronesi, 2009). These complexities cause the future cash flow to evolve in an uncertain and chaotic environment and that is why managing risk is one of the main topics in finance.

An important class of time series models are the so-called long-memory processes which were first introduced by

An important class of time series models are the so-called long memory processes which were first introduced by (Mandelbrot, 1963), (Granger and Joyeux, 1980) and (Hosking, 1981). Long-memory processes have a very useful property which is that their autocorrelations decrease at a slow rate. They are stationary processes that are not quite I(0) and have useful information in them, but are not non-stationary and I(1) series. Their past no matter how far still matter.

Since 1990s with the advancement in technology, data at higher frequencies have become available. Yet, some of the empirical studies use lower frequencies (yearly, monthly) data and test for the long memory process and some use higher frequencies. This availability of the data brings the important question of temporal aggregation and long-memory.

The effects of temporal aggregation on Box and Jenkins' ARIMA models have been extensively studied in the time series literature. Established results such as (Amemiya and Wu, 1972) are mostly in agreement with the fact that the aggregates of an ARIMA(p, d, q) process have a limiting ARIMA(0, d, d) structure which only depend on the parameter d.

With the works of (Chambers, 1998) and (Souza, 2005), the interest to study the effect of systematic sampling (skip sampling) and temporal aggregation on long-memory ARFIMA processes has increased recently. The literature is more focused on the effects of temporal aggregation since it is more relevant for the financial returns with the exceptions of (Chambers, 1998), (HWANG, 2000) and (Souza and Smith, 2002) where they study the effects of systematic sampling on long-memory processes. Through Monte-Carlo studies (Souza and Smith, 2004) and (Man and Tiao, 2006) investigate the effects of temporal aggregation on different ARFIMA(p,d,q) models, while (Souza, 2008) and (Hassler, 2011) consider the spectral properties of temporally aggregated ARFIMA.

There are several estimators used for estimating the long-memory characteristics of the temporally aggregated data, such as the GPH estimator (Geweke and Porter-Hudak, 1983), the Whittle estimator (Whittle, 1953) and other log-periodogram methods. The choices might seem somewhat limited considering there are other available methods for estimating the long-memory characteristics of the data, that are not used in the context of temporal aggregation.

The ultimate goal of this chapter is to establish a framework for modeling the dynamic distribution of financial returns at different risk horizons and to use and compare a more recent estimator, Wavelet Maximum Likelihood Estimator (WMLE) first introduced by (Whitcher, 2000) and compare its performance with those more frequently used, namely the GPH and Whittle estimators. We are the first to use the Wavelet Maximum Likelihood estimator in the context of temporal aggregation.

Chapter 3 is organised as follows. Sections 3.2 and 3.3 describe the data used in this chapter and give the definitions of the asset returns respectively. The dependence properties of financial returns are discussed in section 3.4 and concepts such as long-memory processes, Autoregressive Fractional Moving Average (ARFIMA) are explained and discussed in details. Different methods for testing and estimating ARFIMA models are expounded, namely the R/S test statistic, the GPH estimator, and the Wavelet Maximum Likelihood Estimator. Section 3.5 explores other properties of financial returns and the famous question "What is the relevant horizon for risk management?" is proposed and answered. Section 3.6 starts by exploring concepts of systematic sampling and temporal aggregation followed by identifying the effects of temporal aggregation on ARFIMA(0,d,0) theoretically and through simulations. This section ends by applying the techniques on $S\&P500$ daily, 5-day and 10-day returns.

3.2 Data

The data used in this chapter are the daily closing prices of S&P500 from 1 January 2001 through 31 December 2010 which is obtained from DataStream. The weekend quotes from Friday 21:05 GMT to Sunday 21:00 GMT were removed. Apart from this, there are not any other filterings to the data set nor any other exclusions of the data points. The daily closing prices of S&P500 are then sampled at 5-day and 10-day frequencies and table (3.1) lists a summary of the data used in this chapter.

Symbol	Observations	Start	End
Daily S&P500	2,514	01/01/2001	31/12/2010
5-day S&P500	502	01/01/2001	31/12/2010
10-day S&P500	251	01/01/2001	31/12/2010

Table 3.1: Total number of observation, starting point and the ending point for the closing prices of S&P500 from 1 January 2001 through 31 December 2010 observed ever 1-Day, 5-Day and 10-Day.

3.3 Asset Return Definitions

It is widely known that daily price movements do not show any significant correlations and look like random walks (Malkiel, 2003), (Malkiel and Fama, 1970) and (Pagan, 1996). Because returns have much better statistical properties than price levels, risk modeling focuses on describing the dynamics of returns rather than prices.

The simple rate of return (r_t) from the closing prices of the asset is defined as:

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{3.1}$$

Where P_t is the daily closing price at time t and P_{t-1} is the daily closing price at time $t-1$. The continously compounded rate of return on an asset is defined as[1]:

$$R_t = ln(P_t) - ln(P_{t-1}) \tag{3.2}$$

where $ln(*)$ is the natural logarithm of $*$. The two rates of return are fairly similar, as re-writing equations (3.1) and (3.2) arrive at:

$$R_t = ln(P_t) - ln(P_{t-1}) = ln\left(\frac{P_t}{P_{t-1}}\right) = ln(r_t + 1) \approx r_t \tag{3.3}$$

The approximation holds because using the Taylor series expansion of $ln(1 + r_t)$ when r_t is close to 0:

$$R_t = ln(1 + r_t) = r_t - \frac{{r_t}^2}{2} + \frac{{r_t}^3}{3} - \frac{{r_t}^4}{4} + \dots \tag{3.4}$$

The two definitions of return carry the same information but for our purpose log returns are preferred. Most assets have a lower bound of zero on the price. Log returns

[1]To prove why the continously compounded rate of return can be written as in equation (3.2), please see Appendix A

are more convenient for keeping this lower bound in the risk model as a negative log return tomorrow will still mean a positive price at the end of tomorrow. When using log returns (R_t) from equation (3.2), tomorrow's closing price can be calculated from:

$$P_{t+1} = exp(R_{t+1})P_t \tag{3.5}$$

Estimating the mean of returns using a sample of daily observations:

$$\hat{\mu} = \frac{1}{T}\sum_{t=0}^{T}((ln(P_t) - ln(P_{t-1})) \tag{3.6}$$

$$\hat{\mu} = \tfrac{1}{T}\left[(ln(P_1) - ln(P_0)) + (ln(P_2) - ln(P_1)) + \ldots + (ln(P_{T-2} - ln(P_{T-1})) + (ln(P_T) - ln(P_T))\right]$$

$$\hat{\mu} = \frac{1}{T}(ln(P_T) - ln(P_0)) \tag{3.7}$$

When estimating the mean of returns only the first and the last observations matter and all the intermediate terms cancel out. To estimate the variance using a sample of daily returns we have:

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=0}^{T}((ln(P_t) - ln(P_{t-1})) - \hat{\mu})^2 \tag{3.8}$$

An obvious difference between the sample mean and the sample variance estimators is that the intermediate prices do not cancel out in the variance estimator. Since all the return observations are squared before they are summed in the average, the returns are now important. At short horizons (such as daily) the mean of the returns μ is considerably smaller than the standard deviation σ and can be safely assumed to be zero in comparison to the standard deviation. Equation (3.8) can be written as:

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=0}^{T}(ln(P_t) - ln(P_{t-1}))^2 = \frac{1}{T}\sum_{t=0}^{T}(R_t)^2 \tag{3.9}$$

3.4 Dependence Properties of Daily Returns

The following tendencies or the so-called stylised facts apply to most financial asset returns. The daily closing prices of S&P500 for the period of 1 January 2001 through 31 December 2010 are used to illustrate each of the features.

Daily returns have very little autocorrelation. That is:

$$\gamma(\tau) = Corr(R_t, R_{t-\tau}) \approx 0, \tau = 0, 1, 2, 3, \ldots, T \tag{3.10}$$

In other words, returns are almost impossible to predict from their own past. Correlation measures the linear dependence between two variables and autocorrelation measures the linear dependence between the current value of a time series variable and the past value of the same variable. Autocorrelation is a crucial tool for detecting linear dynamics in time series analysis and it captures the linear relationship between today's value and the value τ days ago[1]. The top part of figure (3.1) shows the daily S&P500 returns accompanied with their autocorreltion functions with returns lagged from 1 to 252 days (roughly a year).

Equation (3.10) is only capable of finding linear dependence between lags of the returns and if there are any non-linear dependence patterns between the present and the past values of the variable, one way to look for them is to use nonlinear transformations of the returns such as squared returns and absolute returns. Variance measured for instance by squared returns, displays positive correlation with its own past. This is obvious from the middle part of figure (3.1) which shows the autocorreltion in daily squared returns, that is:

$$\Gamma_2(\tau) = Corr(R_t^2, R_{t-\tau}^2) > 0, \quad for\ some\ \tau \tag{3.11}$$

It is discussed heavily in the literature (going back to at least (Mandelbrot, 1971), and more recently in (Bollerslev and Wright, 2000), (Cont, 2001)), that linear filtering statistical tools, such as autocorrelation analysis in the time domain analysis, are not capable of distinguishing between asset returns and white noise. Roughly speaking, the absence of autocorrelation in the returns only gives some empirical evidence for random walk models of prices, but it does not mean that the increments (ϵ_t) are

[1]The correlation between the return in the periods of r and s is given by: $corr(R_r, R_s) = \frac{\kappa_R(r,s)}{\sigma_r \sigma_s}$.

also independent. If the increments were to be independent, then any nonlinear functions (e.g. absolute returns, squared returns) of returns should also show no signs of autocorrelation.

If the autocorrelation function of absolute return is defined as:

$$\gamma_{||}(\tau) = corr(|R(t)|, |R(t+\tau)|) \tag{3.12}$$

which is shown in the bottom part of figure (3.1), one can observe that autocorrelation of absolute returns exhibit significant positive autocorrelations. As explained in (Mandelbrot, 1963), *"large price changes are not isolated between periods of slow change but large changes tend to be followed by large changes of either positive or negative signs" ..."small changes tend to be followed by small changes"*. This is why the autocorrelations of absolute returns are significantly positive. This is also in line with the findings of (Bollerslev and Wright, 2000) and (Cont, 2001) and shows that price increments (ϵ_t) are not independently distributed. What is useful here is that, the magnitude (the absolute value) of future returns can be predicted (to some extent) based on past returns, even though the sign of price changes (returns) cannot be forecasted. This result is well known as clustered volatility and has been verified in many financial markets [to name a few: Cont et al. (1997), Bollerslev and Wright (2000), Cont (2001)].

Equation (3.12) gives the autocorrelation for absolute return, however, technically one can study autocorrelations of any nonlinear functions of returns, for example various powers of returns:

$$\gamma_{\alpha}(\tau) = corr(|R(t)|^{\alpha}, |R(t+\tau)|^{\alpha}) \tag{3.13}$$

Or, logarithm of absolute returns, which is rooted in multifractal stochastic volatility models:

$$\gamma_{ln}(\tau) = corr(ln|R(t)|, ln|R(t+\tau)|) \tag{3.14}$$

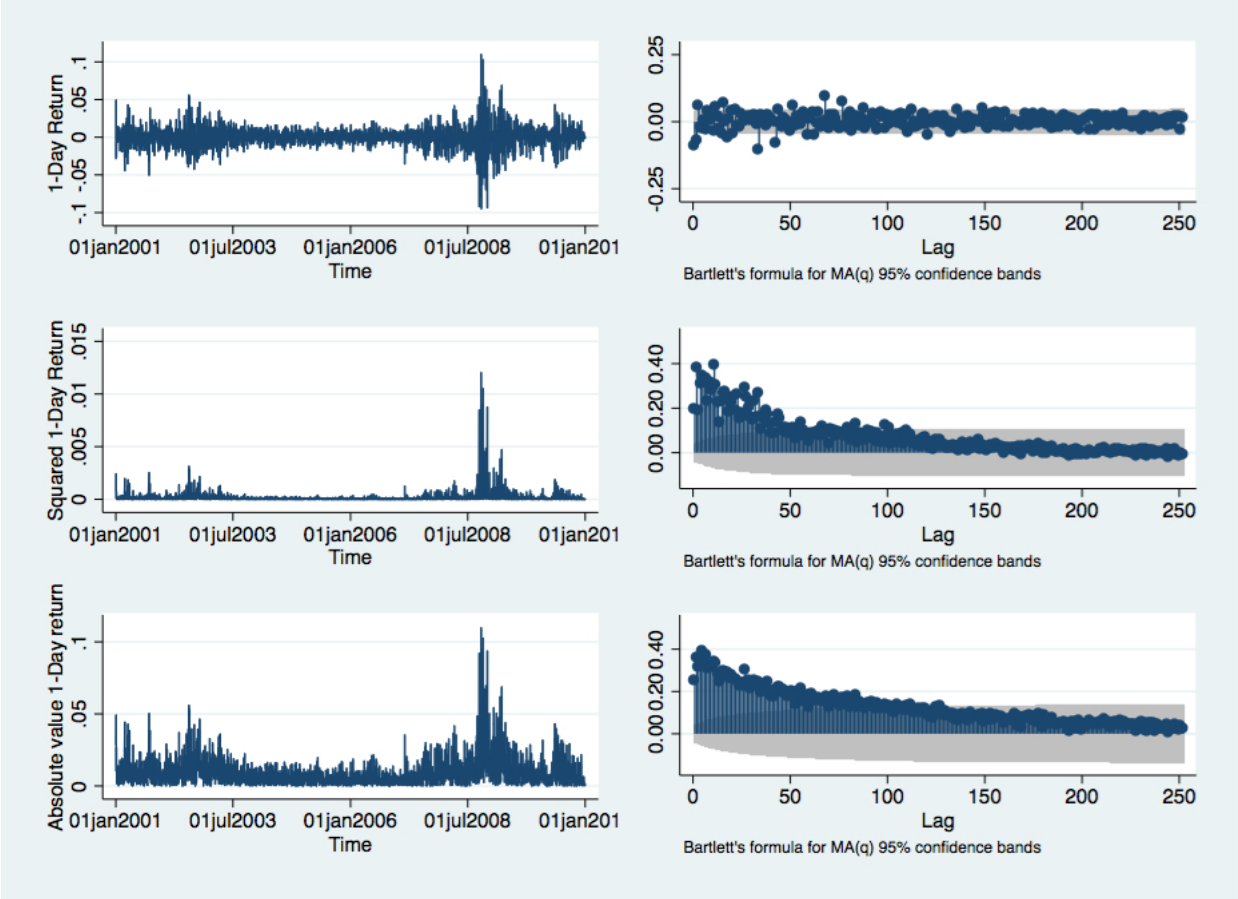

Figure 3.1: The top row of graphs show the daily returns and autocorrelations of the daily returns. The middle rows shows the squared daily returns and their autocorrelations. The bottom row shows the absolute returns and their autocorrelation. The lag order on the horizontal axis refers to the number of days between the return and the lagged return for a particular autocorrelation. It is clear that as we move from top to bottom, autocorrelation becomes more significant and visible. It is obvious that the autocorrelation for the absolute daily returns are the most predictable among these three.

What matters regarding these formulae is that these nonlinear functions of the returns give more insight into the dependence properties of returns[1]. These nonlinear functions of returns which form different *volatility models* show signs of predictability in contrast with return that does not show any signs of predictability, i.e., there is hope in predicting *volatility* rather than the returns themselves. To what extent or how far into the future these non-linear transformations of returns are predictable is of great importance to risk managers.

Comparing the top and the bottom parts of figure (3.1), it is easy to see that the degree of predictability of these variables are different. As we move from top to bottom, there are more significant positive autocorrelations. In other words, what has happened far into the past (100 days ago) matters to the current positions of squared and absolute daily returns, whereas, for the current values of the return series only few lagged returns are useful. In the literature if a series has only few significant autocorrelations is called a short memory process, whereas, if what has happened far into the past matter today it is a long memory process.

3.4.1 Long-Memory Processes

Short-memory processes are stationary processes whose autocorrelation functions converges to zero after few lags. Long-memory processes are stationary processes whose autocorrelation functions decays much more slowly than short-memory processes, but unlike nonstationary processes the autocorrelation functions seem to converge at some point. Since the autocorrelations die out so slowly, long-memory processes are also called long-range dependent or long-run dependent processes.

It might seem to the naked eye that since these autocorrelation functions are all very small numbers, maybe there is no need to be concerned about the ways these variables are treated. To check this hypothesis, it would help to look deeper into the statistical properties of $var(\overline{X})$ (the variance of the sample mean) and observe consequences of not treating a long-memory variable correctly (see (Beran, 1994)).

[1](Ding et al., 1993) show empirically and (Forsberg and Ghysels, 2007) explain theoretically that $\gamma_\alpha(\tau)$ is highest when $\alpha = 1$, meaning that $\gamma_{|\cdot|}(\tau)$ outperforms other powers of returns $\Gamma_\alpha(\tau), \alpha > 1$. That is, absolute returns are more predictable than any other powers of returns.

If $X_1, ..., X_n$ are observations with the same mean $\mu = E(X_i)$ and the variance $\sigma^2 = var(X_i) = E[(X_i - \mu)^2]$, then the variance of $\overline{X} = \frac{\Sigma_{i=1}^{i=n} X_i}{n}$ is:

$$var(\overline{X}) = \frac{\sigma^2}{n} \tag{3.15}$$

There are three conditions under which equation (3.15) holds, which are:

- The population mean $\mu = E(X_i)$ exists and is finite.
- The population variance $\sigma^2 = var(X_i)$ exists and is finite.
- $X_1, ..., X_n$ are uncorrelated, that is: ($\Gamma(r, s) = 0$, for $r \neq s$, where $\Gamma(r, s) = \frac{\gamma(r,s)}{\sigma^2}$, where $\gamma(r, s)$ is the autocovariance between X_r and X_s.

The last assumption is the one that determines whether a variable is short-memory or a long-memory process as it is what matters the most to economists as it makes the future (not) predictable. If observations are correlated, assuming that $\mu = E(X_i)$ is a constant, the variance of ($\overline{X} = \frac{\Sigma_{i=1}^{n}}{X}_i n$) is:

$$var(\overline{X}) = \frac{\sum_{r,s=1}^{n} \gamma(r, s)}{n^2} = \frac{\sigma^2 \Sigma_{r,s=1}^{n} \Gamma(r, s)}{n^2} \tag{3.16}$$

Consider X_t to be an AR(1) process (short-memory):

$$X_t = \alpha X_{t-1} + \epsilon_t, \alpha \in (-1, 1) \tag{3.17}$$

where ϵ_t are normally distributed with zero mean and constant variance. Then:

$$var(\overline{X}) = \frac{\sigma^2 \sum_{r,s=1}^{n} \Gamma(r, s)}{n^2} \tag{3.18}$$

Assuming a covariance stationary process, then:

$$\Gamma(r, s) = \alpha^{|r-s|} = \alpha^{|s-r|} = \alpha^k \tag{3.19}$$

Which then simplifies to:

$$var(\overline{X}) = \frac{\sigma^2 \left[\sum_{i=1}^{n} 1 + \sum_{r \neq s} \alpha^{|r-s|}\right]}{n^2} \tag{3.20}$$

Using the rule of geometric sums:

$$\sum_{k=1}^{n-1} \alpha^k = \frac{\alpha - \alpha^n}{1-\alpha} \tag{3.21}$$

Equation (3.20) can be written as:

$$var(\overline{X}) = \frac{\sigma^2}{n}\left[1 + \left[\frac{2\alpha}{1-\alpha}(1 - \frac{1}{n(1-\alpha)} + \frac{\alpha^n}{n(1-\alpha)})\right]\right] \tag{3.22}$$

and for large enough n:

$$var(\overline{X}) = \frac{\sigma^2}{n}\left[1 + \left[\frac{2\alpha}{1-\alpha}\right]\right] = \frac{\sigma^2}{n} \times \frac{1+\alpha}{1-\alpha} \tag{3.23}$$

Now, consider the case when α is close to ± 1. If $\alpha \to 1$, then $var(\overline{X}) \to \infty$ and if $\alpha \to -1$, then $var(\overline{X}) \to 0$. In such cases the equation (3.15) is not a good approximation of the variance of $\overline{X}$. Equation (3.23) can be re-written as:

$$var(\overline{X}) = \frac{\sigma^2}{n} A(\alpha) \tag{3.24}$$

Equation (3.24) is different from equation (3.15) in the sense that it allows $A(\alpha)$ to differ from 1. One might think that equation (3.24) is general enough and is able to capture the characteristics of all data types. It turns out that for some data sets in economics, finance,..., etc, that is not the case and the variance of $\overline{X}$ is different from equation 3.15 not just by $A(\alpha)$ but by the speed at which it converges to zero (look at the top and the bottom part of figure (3.1). If the variance of $\overline{X}$ converges to zero slower than $\frac{1}{n}$, then $A(\alpha)$ is not a constant function of α but it is a function that increases with n. Fitting the best ARMA model will then lead to an ARMA model with *many* parameters. As sample size increases, the number of parameters will tend to infinity. Using an excessive number of parameters is not desirable, since it increases the uncertainty of the statistical power, and the parameters will be difficult to interpret (Sowell, 1992b). That is why long-memory processes will not be best analysed using the models that work for short-memory processes.

The detection of long-range dependence is not very easy and can't be done by only looking at one or a few selected correlations. We rather have to judge the way the correlations converge to zero with increasing lag. It is the statistical power of all

correlations combined that determines whether dependence is rejected or accepted at a reasonable statistical power. The literature on long-memory time series suggests to treat the slow decaying process as being proportional to n^{-a} for $a \in (0,1)$, rather than to $\frac{1}{n}$. This makes equation (3.26):

$$var(\overline{X}) \approx \frac{\sigma^2}{n^a} A(\alpha) \tag{3.25}$$

where, $a \in (0,1)$ is a constant and $A(\alpha)$ is defined as:

$$A(\alpha) = \lim_{n\to\infty} n^{\alpha-2} \sum_{r\neq s} \Gamma(r,s) \tag{3.26}$$

where $\sum_{r\neq s} \Gamma(r,s)$ is not summable as (Beran, 1994) prove that $\Gamma_\tau \approx C_1 \times \tau^{-a}$ for $a \in (0,1)$ and a constant C_1[1]. A process is said to have long memory if the dependence between events that are still far apart from each other lessens very slowly as distance increases, that is $\Gamma_\tau \approx |\tau|^{-a}$. It is not important whether some Γ_τ s are smal for some τ s to start with, but it matters if the rate at which the correlations decrease is *very* slow. Recall from chapter two that power spectrum $\hat{f}(\nu)$ is defined as:

$$\hat{f}(\nu) = \frac{1}{2\pi} \sum_{-n}^{n} \Gamma(\tau) e^{-i\nu\tau} \tag{3.28}$$

Assuming $\Gamma_\tau \approx C_1 \times |\tau|^{-a}$ if $\nu \to 0$, then:

$$\hat{f}(\nu) \approx \frac{1}{2\pi} \sum_{-n}^{n} \Gamma(\tau) e^{0} \tag{3.29}$$

And hence:

$$\hat{f}(\nu) \approx C_2 |\tau|^{a-1} \tag{3.30}$$

Which means $\hat{f}$ goes to infinity, near the origin (when frequency ($\nu \to 0$)) if $d-1 < 0$ or $d < 1$. Hence another way to spot long-memory processes, is to look at their spectral representations. If the spectrum goes to ∞ near the origin ($\nu \to 0$) then there is definite

[1]
$$\sum_{k=-\infty}^{\infty} \Gamma_\tau = \sum_{k=-\infty}^{\infty} \frac{C_1}{|\tau|^d} = \infty \tag{3.27}$$

Equation 3.27 is a result of the fact that, $\sum_{x=1}^{\infty} \frac{1}{x^y}$ diverges if $y \leq 1$.

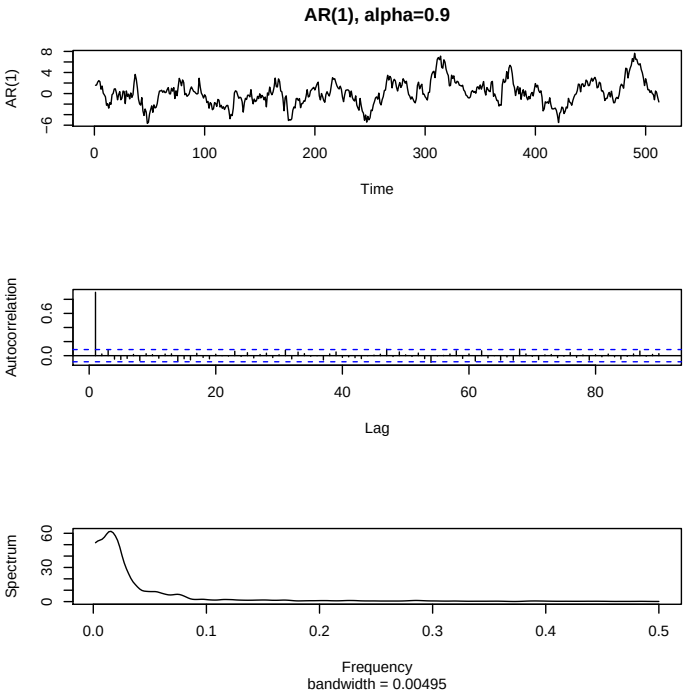

Figure 3.2: Plot of simulated AR(1) with $\alpha = 0.9$, against time, partial autocorrelation functions against lag, and power spectrum against frequency.

doubt that the series has long-run dependence. Figure (3.2) represent AR(1) with $\alpha = 0.9$ as an example for a short memory process. It is clear from the power spectrum that this is a short-memory process since near the origin, the spectrum doesn't go to infinity. It goes quite high but then levels off which is because of the choice of $\alpha = 0.9$.

Defining Long-Memory

Let $\gamma(k)$ be the autocovariance function at lag k of the stationary process x_t. A usual definition of long memory is that:

$$\sum_{k=-\infty}^{\infty} |\gamma(k)| = \infty \tag{3.31}$$

However, there are other alternative definitions used in the literature. Using the slow hyperbolic decay rate of the autcovariances:

$$\gamma(k) \approx k^{2d-1} f(k) \tag{3.32}$$

as $k \to \infty$, where d is the so-called long-memory parameter and $f(*)$ is a slowly varying function. Making use of the spectral representation another definition is:

$$g(\nu) \approx |\nu|^{-2d} h\left(\frac{1}{|\nu|}\right) \tag{3.33}$$

as $\nu \to 0$ and $h(*)$ is slowly varying function. There are other alternative definitions of long-memory processes which are not necessarily equivalent to each other. In this thesis the two definitions in equations (3.32) and (3.33) are frequently used.

3.4.2 Autoregressive Fractional Integrated Moving Average (ARFIMA)

A well known class of short-memory models is the Autoregressive Integrated Moving Average (hereafter abbreviated to ARIMA(p,d,q)) model is a linear univariate time series model which depends on three parameters:

$$\phi(L)(1-L)^d x_t = \delta(L)\epsilon_t \tag{3.34}$$

Where the first set of p parameters in $\phi(L)$ define the autoregressive polynomial in the lag operator L and the second q parameters in $\epsilon(L)$ represent the moving average polynomial in the disturbance process:

$$\phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$$
$$\delta(L) = 1 - \delta_1 L - ... - \delta_q L^q$$

The third parameter d, expresses the integer order of differencing that should be applied to the series before estimation. For an $ARIMA(p, d, q)$ model, there are $p + q$ parameters that need to be estimated. When estimating an ARIMA model, the integer order of differencing d is chosen to ensure that $(1 - L)^d x_t$ is a stationary process, hence often given the name $I(d)$ processes.

ARIMA(p,d,q) with an integer d is widely used for modeling short memory processes where the autocorrelation functions dampens after few lags. Many time series happen to exhibit too much long-range dependence to be classified as $I(0)$ but also they are not $I(1)$, simply the memories of the past stay with the process for much longer than in stationary ARMA processes (see the middle and bottom panels of figure (3.1))[1]. The family of ARIMA processes can be generalized by permitting the degree of differencing to take fractional values. The fractional differencing operator is defined as an infinite binomial series expansion in powers of the backward-shift operator. (Hosking, 1981) and (Granger and Joyeux, 1980) suggested that non-integer values of d can be useful, and introduced Autoregressive Fractionally Integrated Moving-Average (hereafter abbreviated to ARFIMA) models, which can be thought of as a generalization of the standard $ARIMA(p, d, q)$ models defined by Box and Jenkins (1994).

Among others (Granger and Joyeux, 1980) and (Box et al., 2008) show that the autocorrelations from an ARMA model decay exponentially, whereas, the autocorrelations from an ARFIMA model decay at much slower hyperbolic rate. A number of studies argue that the long-run dependence in financial market volatility can be

[1]To be precise: (Granger and Joyeux, 1980) argue that *if a series does need differencing, it means that strictly the original, undifferenced series has infinite variance. ... This has led time series analysts to suggest that econometricians should at least consider differencing their variables when building models. However, econometricians have been somewhat reluctant to accept this advice, believing that they may lose something of importance. Phrases such as differencing "zapping out the low frequency components" are used.* They clearly are arguing that taking the first difference of a time series may not necessarily be the best remedy when the series is between $I(0)$ and $I(1)$. (Palma, 2007) and (Robinson, 2003) also point out the fact that over-reliance on Dickey-Fuller tests which have $I(1)$ as the null hypothesis and $I(0)$ as the alternative, can result in disappointing verdicts.

conveniently modeled by a fractionally-integrated process, so that although volatility shocks are highly persistent, they eventually dissipate at a slow hyperbolic rate [Hosking (1981), Pong et al. (2004), Lux and Kaizoji (2007), Corsi (2009)]. ARFIMA models are designed to be flexible enough to represent series that are between $I(0)$ and $I(1)$ and to accommodate the representation of long-memory processes. Any ARMA stationary time series is classified as short-memory time series. All the past memories of an $ARIMA(p, d, q)$ model are embedded in x_t and ϵ_t, but the effect of past values dampens off exponentially to near-zero values quickly. ARFIMA models are used for when the the time series has a long memory. Long-memory processes were first introduced by (Hurst, 1951) and (Hosking, 1981) in hydrology, and (Granger and Joyeux, 1980) in economics independently discovered the ARFIMA representation of long-run dependence. (Beran, 1994), (Baillie, 1996), (Palma, 2007) and (Robinson, 2003) provide good introductions to long-memory processes and ARFIMA models.

The usual differencing procedure consists of using the operator $(1 - L)$, where L is the backward-shift operator. For example $(1 - L)^2 = 1 - 2L + L^2$ or $(1 - L)^3 = 1 - 3L + 3L^2 - L^3$, where $Lx_t = x_{t-1}$, $L^2x_t = x_{t-2}$, and $L^3x_t = x_{t-3}$. In fractional model, the power is allowed to be fractional, such as in the binomial series:

$$\text{(1-L)}^d = \sum_{i=0}^{\infty} \binom{d}{i} (-L)^i$$
$$\text{(1-L)}^d = 1 - dL + \frac{d(d-1)}{2!} L^2 - \dots$$

Therefore

$$\text{(1-L)}^d x_t = x_t - dx_{t-1} + \frac{d(d-1)}{2!} x_{t-2} + \dots$$

Suppose there is a filter $F(L)$ such that when used twice, one gets the usual difference, i.e., $F^2(L) = (1 - L)$. According to Granger (1980) such a filter exists and if this filter is used once, it can be thought of as "half differencing", which is an example of fractional differencing with $d = 1/2$. A fractional $ARIMA(p, d, q)$ with nonnegative integers p and q and real d is a time stochastic process which can be represented as:

$$\phi(L)(1 - L)^d x_t = \delta(L)\epsilon_t \tag{3.35}$$

$$\phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$$
$$\delta(L) = 1 - \delta_1 L - ... - \delta_q L^q$$

Where $E(x_t) = 0$ and the $\{\epsilon_t\}$ are independent and identically distributed random variables with mean 0 and variance σ_ϵ^2. An $ARFIMA(p, d, q)$ process is both stationary and invertible if all roots of $\phi(L)$ and $\delta(L)$ lie outside the unit circle and $|d| < 1/2$. (Hosking, 1981) explains that the effect of the d parameter on distant observations decays hyperbolically as the lag increases, while the effects of the ϕ and δ parameters decay exponentially. Thus d may be chosen to describe the high-lag correlation structure of a time series while ϕ and δ parameters are chosen to describe the low-lag correlation structure. Indeed the long term behavior of an $ARIMA(p, d, q)$ process may be expected to be similar to that of an $ARIMA(0, d, 0)$ process with the same value of d, since for very distant observations the effects of the ϕ and δ parameters will be negligible. An ARFIMA process defined by equation (3.35) is stationary for values of $-0.5 < d > 0.5$ (Granger and Joyeux, 1980).

The autocovariance (κ), autocorrelation (γ) and spectral density functions ($\hat{f}(\nu)$) of ARFIMA(0,d,0) for $\tau \geq 0$ and $0 < \nu \leq \pi$ are:

$$\begin{aligned} \kappa_\tau &= \frac{\Gamma(1-2d)\Gamma(\tau+d)}{\Gamma(d)\Gamma(1-d)\Gamma(\tau+1-d)}\sigma_\epsilon^2 \\ &\sim \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}\tau 2d-1\sigma_\epsilon^2, \ \ as \ \ \tau \longrightarrow \infty \end{aligned} \tag{3.36}$$

$$\begin{aligned} \gamma_\tau &= \frac{\Gamma(1-d)\Gamma(\tau+d)}{\Gamma(d)\Gamma(\tau+1-d)} \\ &\sim \frac{\Gamma(1-d)}{\Gamma(d)}\tau^{2d-1}, \ \ as \ \ \tau \longrightarrow \infty \end{aligned} \tag{3.37}$$

$$\begin{aligned} \hat{f}(\nu) &= \frac{\sigma_\epsilon^2}{2\pi}(2sin(\tfrac{\nu}{2})^{-2d} \\ &\sim \frac{\sigma_\epsilon^2}{2\pi}\nu^{-2d}, \ \ as \ \ \nu \longrightarrow 0 \end{aligned} \tag{3.38}$$

Where $d \in (-1/2, 1/2)$, Γ is the gamma function and Stirling's approximation ($\frac{\Gamma(\tau+a)}{\Gamma(\tau+b)} \sim \tau^{a-b}$) has been used.

Proof: See the Appendix.

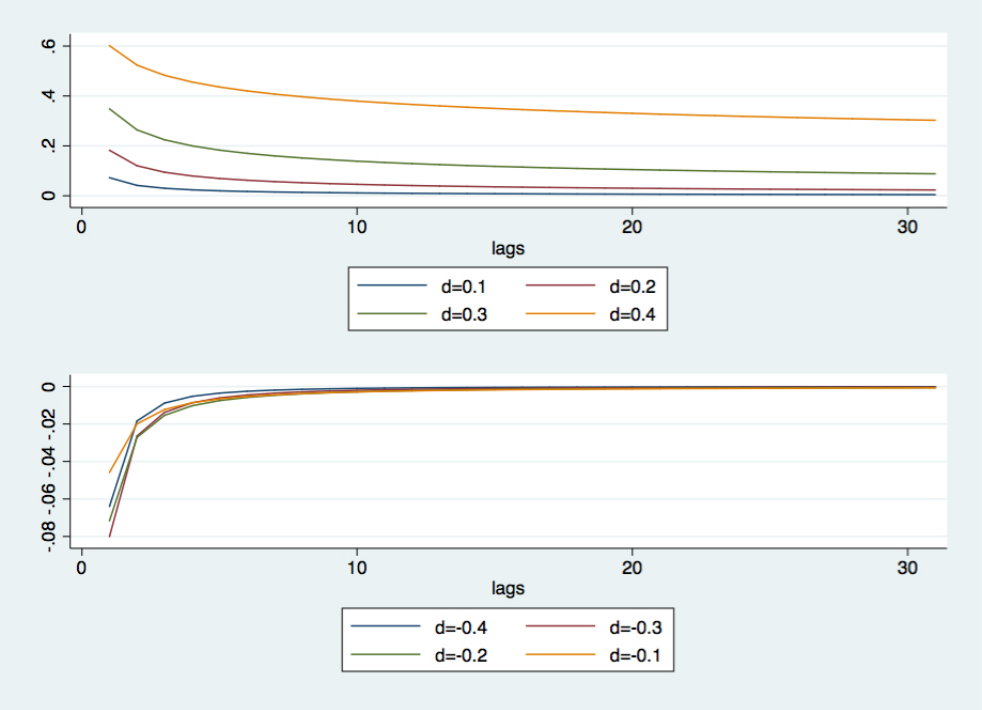

Figure 3.3: Autocorrelation coefficients for an ARFIMA(0,d,0) for different values of d, positive values of d in the top graph and negative values of d in the bottom graph.

Figure (3.3) represents the slow decaying rates of autocorrelation functions (equation (3.37)) of ARFIMA(0,d,0) processes for different values of d. (Hosking, 1981) show that $\hat{f}\nu$, the spectral density function and $\gamma\tau$, the autocorrelation function for ARFIMA(p,d,q) processes also have these properties:

$\lim \nu^{2d}\hat{f}(\nu)$ exists, as $\nu \longrightarrow 0$, and is finite, i.e. $\hat{f}(\nu) \propto \nu^{-2d}$ approximately as $\nu \longrightarrow 0$.

$\lim \tau^{1-2d}\gamma_\tau$ exists, as $\tau \longrightarrow \infty$, and is finite, i.e. the correlation function of an ARFIMA process is proportional to τ^{2d-1} as $\tau \to \infty$.

The d in the above theorems is exactly analogous to a in equations 3.30 and 3.26; $2d = a$. If $d \in (-1/2, 0)$ the process is said to exhibit intermediate memory (anti-persistence), or long-range negative dependence (Baum et al., 1999). The process exhibits short memory for $d = 0$, corresponding to stationary ARIMA modeling and if $d \in [1/2, 1)$ the process is mean reverting but it is not stationary. If $d \in (0, 1/2)$ the ARFIMA process is said to exhibit long memory or long-run positive dependence.

3.4.2.1 ARFIMA Models, Testing and Estimating

There are different approaches to the estimation of an ARFIMA(p,d,q) model which can be classified into two groups; parametric and semiparametric methods. The most popular methods within the first group are (Sowell, 1992a) and (Fox and Taqqu, 1986) which involve the likelihood function. In the latter group, the most commonly used methods are the GPH method which was proposed by (Geweke and Porter-Hudak, 1983) and the modified rescaled range R/S test proposed by (Lo, 1991) which was a modification of the classical R/S statistic initially proposed by (Hurst, 1951) and (Mandelbrot, 1972). In semiparametric methods, it is assumed that the short-memory components of the time series are relatively less important and the focus is on estimating the parameter d, whereas, in the parametric approaches choosing an appropriate ARMA representation is always a challenge. To model the long-run feature of the daily S&P500 returns, a combination of methods are used and explained as each is complementary to the other.

The Modified Rescaled Range R/S, (Lo, 1991)

This semiparametric test is rooted from the original R/S test ((Hurst, 1951)), is the range of the partial sums of deviations of a time series from its mean, then rescaled by its standard deviation. For example for a sample of n values $X_1, ..., X_n$, the test statistic is:

$$Q_n = \frac{1}{S_n}[Max_{1\leq k\leq n}\sum_{j=1}^{k}(x_j - \bar{x_n}) - Min_{1\leq k\leq n}\sum_{j=1}^{k}(x_j - \bar{x_n})] \tag{3.39}$$

where S_n is the maximum likelihood estimator of the standard deviation of x. $Max_{1\leq k\leq n}\sum_{j=1}^{k}(x_j - \bar{x_n})$ is the maximum of the partial sums of the first k deviations from the mean, which is non-negative. $Min_{1\leq k\leq n}\sum_{j=1}^{k}(x_j - \bar{x_n})$ is the minimum of the partial sums of the first k deviations from the mean, which is non-positive. The difference of the two quantities will be such that $Q_n > 0$. (Hurst, 1951) and (Lo, 1991) demonstrated that the R/S statistic is able to detect long-range dependence in the data. Like all the other long-range dependence estimators, the R/S statistic is excessively sensitive to short-term dependence. (Lo, 1991) shows that an AR(1) process with large sample size can seriously bias the R/S statistic. He then modified the test to account for the short-term effect by applying a "Newey-West" correction (using a Bartlett window) to derive a consistent estimate of the long-range variance of the time series. S is replaced by $\hat{S}$:

$$\hat{S} = \sqrt{S^2 + 2\sum_{\tau=1}^{r}(1 - \frac{\tau}{r+1})\Gamma(\tau)} \tag{3.40}$$

wher r is the maximum lag over which short-term autocorrelation might be substantial and $\Gamma(\tau)$ is the autocorrelation at lag τ. If the maximum lag r is set to zero the, the statistic falls back to what was first proposed by (Hurst, 1951). We can test whether daily S&P500 returns are predictable and have long memory (i.e., to test if EMH holds) using the original Hurst Statistics (Hurst, 1951). Setting $r = 0$ in equation (3.40) will result in the original Hurst statistic:

```
. lomodrs  dayreturn, max(0)

Hurst Classical R/S test for dayreturn

Critical values for H0: dayreturn is not long-range dependent
```

	Return	Squared Return	Absolute Return
1-Day S&P500	1.19	2.94	3.83

Table 3.2: The summary of the Modified R/S test statistic for the return, squared return and absolute return at 1-Day, 5-Day and 10-Day frequencies. The Critical Values are as follows: 90% : $[0.861, 1.747]$, 95% : $[0.809, 1.862]$, 99% : $[0.721, 2.098]$

```
90%: [ 0.861, 1.747 ]
95%: [ 0.809, 1.862 ]
99%: [ 0.721, 2.098 ]

Test statistic:     1.19    N = 2513
```

Which gives a statistic of 1.19. The H_0: the variable is not long-range dependent and H_1: the variable is long-range dependent. If the test statistic is within the critical values boundaries, then H_0 is not rejected. In this case 1.19 falls withing all three intervals and hence H_0 is accepted and there is no evidence that the daily returns are long-range dependent. Using the modifies version of (Lo, 1991) statistic, assuming that the last 10 days contains the short-term dependence[1]:

```
. lomodrs  dayreturn, max(10)

Lo Modified R/S test for dayreturn

Critical values for H0: dayreturn is not long-range dependent

90%: [ 0.861, 1.747 ]
95%: [ 0.809, 1.862 ]
99%: [ 0.721, 2.098 ]

Test statistic:     1.38  (10 lags)  N = 2503
```

As the test statistic (1.38) is between the critical values at 90%, 95% and 99% , there is no evidence to reject the H_0. Table (3.2) summarises the Modified R/S test for squared returns and absolute returns which shows evidence that H_0 is rejected at

[1]Without specifying what the maximum lag is the test statistic is 1.19 which still accepts the H_0. Critical values are taken from (Lo, 1991), Table II

all levels and there is evidence of long-run dependence in the daily squared and daily absolute returns.

The Geweke-Porter-Hudak Log Periodogram Regression Estimator(Geweke and Porter-Hudak, 1983)

The Geweke and Porter-Hudak method often known as GPH method, uses a spectral regression estimator to estimate d without any specifications of the short-memory parameters of the series. The proposed semiparametric method obtains an estimate of the memory parameter d for x_t in $(1-L)^d x_t = \epsilon_t$. If the periodogram for x_t which is the squared modulus of the discrete Fourier transform is defined as:

$$I_x(\nu_s) = \frac{1}{2\pi N}|\sum_{t=1}^{N} X_t exp(it\nu_s)|^2, \nu_s = \frac{2\pi s}{n}, s = 1, ...m, m < n \tag{3.41}$$

Which after taking logs from both sides of the equation 3.41 and some algebra it simplifies to:

$$log(I_x(\nu_s)) = \hat{c} - \hat{d} log\,|1 - exp(it\nu_s)|^2 + e_t \tag{3.42}$$

Where e_t is the residual and $\hat{d}$ can be estimated by applying an ordinary least squares regression to equation (3.42) which gives:

$$\hat{d} = 0.5\frac{\sum_{s=1}^{m} x_s log(I_x(\nu_s)}{\sum_{s=1}^{m} x_s^2} \tag{3.43}$$

The choice of m, the number of Fourier frequencies included in the regression is very crucial to the estimate of parameter d and there are different suggestions in the literature. Basically the slope in equation (3.42) is an estimate of the slope of the series' power spectrum in the neighborhood of the zero frequency. If m is small, then the slope is estimated from a small sample size and if a large m is chosen then, the medium and high-frequency components of the spectrum will also be included in the regression which could corrupt the estimate. The long memory parameter d can be estimated with $s = 1, 2, \ldots, g(N)$, where $lim_{N\to\infty} = \infty$ and $lim_{N\to\infty \frac{g(N)}{N}=0}$. That is, g(N) has to be reasonably small enough compared to N. Often as the default $m = 0.5$ ($g(N) = N^{0.5}$) is chosen. To overcome this problem a range of different m (different

fractions of the original sample size) is often used. The GPH method as it stands does not consider the case $d = 1$ (unit root) and (Phillips, 1999) modified the GPH estimator to be able to accommodate the $d = 1$ case.

(Robinson, 1995) proposed an alternative log-periodogram regression estimator in which the estimator is not restricted to using a small fraction of the sample size and it is more flexible. Robinson's method uses the average of the periodogram over adjacent frequencies rather than using the peridogram of only one frequency. The Robinson's method can be used to test whether the daily S&P500 returns have long memory (another EMH test) or whether they are stationary ($d = 0$):

```
roblpr  dayreturn , j(2) power(0.1 0.15:0.9)

Robinson estimates of fractional differencing parameter for dayreturn
                                Averaging    =      2
-------------------------------------------------------------
Power   Ords      Est d    Std Err  t(H0: d=0)    P>|t|
-------------------------------------------------------------
 .1      2   -.952711          0          .         .
.15      2   -.952711          0          .         .
 .2      3  -.4989558   .4072273    -1.2253     0.266
.25      4  -.2839519     .30003    -0.9464     0.372
 .3      6  -.1704466   .1907618    -0.8935     0.389
.35      8  -.0776705   .1473404    -0.5272     0.605
 .4     12   .0190295   .1002669     0.1898     0.851
.45     17  -.0890828   .0831622    -1.0712     0.292
 .5     26  -.0185011   .0812664    -0.2277     0.821
.55     38   .0595285   .0753749     0.7898     0.432
 .6     55   .0025476   .0587598     0.0434     0.965
.65     82   .0323456   .0475755     0.6799     0.498
 .7    120   -.019882   .0368126    -0.5401     0.590
.75    178  -.0420046   .0297958    -1.4098     0.159
 .8    263  -.0233877    .025285    -0.9250     0.355
.85    389  -.0429392   .0216752    -1.9810     0.048
 .9    575  -.0655842   .0177768    -3.6893     0.000
```

For different m (powers) starting from 0.1 through 0.9 with a 0.15 gap, the hypothesis that $H_0 : d = 0$ is tested against different estimates of d, averaging the log periodogram of the two adjacent frequencies ($j = 2$). It should be noted that "power" in the GPH test refers to the parameter in $g(N)$ and shouldn't be mistaken with power

(sensitivity) of a test. For all powers (except 0.85 and 0.9) H_0 is accepted that the series is stationary and there is no evidence of long-memory ($d = 0$). As mentioned before a large m shouldn't be considered (like 0.85 and 0.9) as the medium and high-frequency components of the spectrum will also be included in the regression, whereas, what we are interested in is the spectrum at low frequencies. If $m = 0.9$ is considered then the H_0 is rejected in favour of $d = -0.066$, which could be interpreted that the daily S&P500 returns are over differenced since d is statistically negative. Testing whether the squared S&P500 daily returns are stationary or not:

```
roblpr  squareddayreturn , j(2) power(0.1 0.15:0.9)
Robinson estimates of fractional differencing parameter for squareddayreturn
                                  Averaging    =      2
-------------------------------------------------------------
Power   Ords      Est d    Std Err  t(H0: d=0)    P>|t|
-------------------------------------------------------------
   .1      2   .5915132          0           .         .
  .15      2   .5915132          0           .         .
   .2      3   .5707397   .0186434    30.6135     0.000
  .25      4   .5755442   .0112922    50.9682     0.000
   .3      6   .5393668   .0400483    13.4679     0.000
  .35      8    .560821   .0488907    11.4709     0.000
   .4     12   .5307021   .0381782    13.9007     0.000
  .45     17   .5130171   .0336406    15.2499     0.000
   .5     26    .756326   .0795921     9.5025     0.000
  .55     38   .6677927   .0616839    10.8260     0.000
   .6     55   .6500147   .0454737    14.2943     0.000
  .65     82   .7399583   .0465813    15.8853     0.000
   .7    120   .5959565   .0433011    13.7631     0.000
  .75    178   .4699798   .0394147    11.9240     0.000
   .8    263   .3104313   .0363845     8.5320     0.000
  .85    389   .2606585   .0276445     9.4290     0.000
   .9    575   .1336727   .0230774     5.7924     0.000
```

The results from Robinson's method are in agreement with the modified rescaled test of (Lo, 1991) that the daily squared returns are not stationary, short-memory processes as $H_0 : d = 0$ is rejected at all powers. The semi-parametric methods ignore the short-memory feature of the time series and hence give estimates of d that are biased upwards. But they are powerful tests for checking whether time series are stationary

short-memory processes or not. If there is evidence of long-memory then the next natural step would be to estimate d more precisely.

Maximum Likelihood Estimators of ARFIMA

The interpretation of ARFIMA models will be facilitated if the short-run and long-run parameters are individually estimated. For ARFIMA models:

$$(1\text{-}\phi_1 L - ... - \phi_p L^p)(1-L)^d x_N = (1 - \delta_1 L - ... - \delta_q L^q)\epsilon_N$$

$$\epsilon_N \sim NID[0, \sigma_\epsilon^2] \tag{3.44}$$

For equation (3.44), the likelihood function is:

$$L = \frac{1}{(2\pi)^{N/2}|\sum|^{1/2}} exp\left(\frac{-X_N' \Sigma^{-1} X}{2}\right) \tag{3.45}$$

where if κ_N is the autocovariance, then the variance matrix of $x = (x_1, \ldots, x_N)'$ is:

$$\sum = VAR(x) = \begin{bmatrix} \kappa_0 & \kappa_1 & \ldots & \kappa_{N-1} \\ \kappa_1 & \kappa_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \kappa_1 \\ \kappa_{N-1} & \ldots & \kappa_1 & \kappa_0 \end{bmatrix}$$

which is a symmetric Toeplitz matrix. The log-likelihood function is:

$$log(L, \delta, \phi, \sigma_\epsilon^2) = -\frac{N}{2} log(2\pi) - \frac{1}{2} log|\sum| - \frac{1}{2} x' \Sigma^{-1} x \tag{3.46}$$

Notably, there are two areas of concerns regarding equation 3.46. First is the evaluation of the autocovariance function and second is handling the $\sum$ matrix which is of dimension $N \times N$.

There are various techniques to compute the autocovariance function of ARFIMA processes. (Bloomfield, 2000) and (Hosking, 1996) used the frequency representation of the autocovariance, and (Chan and Palma, 1998), (Ravishanker, 1997) used an $MA(\infty)$ representation of the time series but the drawback is that there are many parameters that need to be estimated. Among others (Palma, 2007), (Golub, 1996), (Hosking, 1981) and (Sowell, 1992b) expressed γ_i as a function of the autocovariances of $y_t \sim$

$ARFIMA(0,d,0)$ and $z_t(p,q)$ processes. That is, $\gamma_i = \sum_{j=1}^{N} \gamma_j^z \gamma_{i-j}^y$. (Sowell, 1992b) was successful in speeding up the algorithm for calculation the autocovariance functions by evaluation the hypergeometric functions recursively[1].

If R is defined as the matrix of the autocovariances scaled by the error variance:

$$\text{R}= \begin{bmatrix} \kappa_0/\sigma_\epsilon^2 & \kappa_1/\sigma_\epsilon^2 & \dots & \kappa_{N-1}/\sigma_\epsilon^2 \\ \kappa_1/\sigma_\epsilon^2 & \kappa_0/\sigma_\epsilon^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \kappa_1/\sigma_\epsilon^2 \\ \kappa_{N-1}/\sigma_\epsilon^2 & \dots & \kappa_1/\sigma_\epsilon^2 & \kappa_0/\sigma_\epsilon^2 \end{bmatrix}$$

Then $R = \Sigma/\sigma_\epsilon^2$ and replacing $\sum$ in equation 3.46 we have:

$$log(L,\delta,\phi,\sigma_\epsilon^2) = -\frac{N}{2}log(2\pi) - \frac{1}{2}log|R| - \frac{N}{2}log\sigma_\epsilon^2 - \frac{1}{2\sigma_\epsilon^2}x'R^{-1}x \tag{3.47}$$

Differentiating with respect to σ_ϵ^2, setting to zero and solving yields:

$$\sigma_\epsilon^2 = N^{-1}x'R^{-1}x \tag{3.48}$$

and replacing 3.48 in 3.47:

$$\begin{gathered} \log(\text{L},\delta,\phi) = \\ \text{-N/2 } (\log(2\pi)) - \tfrac{N}{2} - \tfrac{1}{2}log|R| - \tfrac{N}{2}(log(N^{-1}x'R^{-1}x) = \\ \text{-N/2}(1+\log 2\pi) - \tfrac{1}{2}log|R| - \tfrac{N}{2}log(N^{-1}x'R^{-1}x) \end{gathered}$$

And finally what is left for the maximisation is:

$$-\frac{1}{2}(log|R| + T^{-1}log\sigma_\epsilon^2) \tag{3.49}$$

which is then maximised using numerical derivatives. Using the same method first introduced in (Sowell, 1992b) and then used for example in (Palma, 2007) and (Golub, 1996), we estimate the parameters of an ARFIMA(0,d,0) for the daily S&P500 squared returns[2]:

[1] for the details in calculating the autocovariances see(Sowell, 1992b) and (Doornik and Ooms, 2003a)

[2] Recall that the daily S&P500 returns did not show any signs of positive persistence.

	Return	Squared Return	Absolute Return
Daily S&P500	-0.0735(-5.04^{***})	0.1952(17.31^{***})	0.2068(18.82^{***})

Table 3.3: The summary of the estimated coefficient for d for the daily returns, daily squared returns and daily absolute returns in ARFIMA(0,d,0). The numbers in the brackets are the z statistics. *** significant at 99% level.

$$(1-L)^{0.1952} r_t^2 = \epsilon_t, \epsilon_t \sim NID[0, 4.02 \times e^{-9}] \tag{3.50}$$

Table (3.3) summarises the estimated d for the daily, squared and absolute returns in ARFIMA(0,d,0). Equation 3.50 is ARFIMA(0,0.3565,0) estimated for the daily S&P500 squared returns. The estimated parameter is highly significant, different from zero and since it is 0.1952, there is evidence of positive persistence in the time series (see figure (3.3)). As discussed before, the spectral density at frequencies close to zero for a long-run processes diverges. Figure (3.4) shows the spectral density for the long-run and the short-run parameters implied by ARFIMA(0,0.3565,0). The difference in the spectral densities in the top graph reflects the ability of ARFIMA models to capture long-run effects.

The Approximate Maximum Likelihood Estimator

The calculation of the exact maximum likelihood estimator in the way proposed by (Sowell, 1992b), when the length of the time series is a large number, is computationally inefficient since it has a numerical order of $O(n^3)$. It is possible to use algorithms such as the Levinsons or the Durbin's algorithm and bring the cost down to $O(n^2)$. Among others, (Doornik and Ooms, 2003b) list and compare the effectiveness of different algorithms for calculating the maximum likelihood, and after extensive simulations they find the combination of Durbin's and Levinson's algorithms to be very efficient. There are however other approaches that are computationally more efficient among which the semiparametric Whittle estimation (Whittle, 1953) is a well-known and widely used methodology for approximating the maximum likelihood (numerical order of $O(n \times logn)$.

Another solution for approximating the maximum likelihood which has recently become popular is the Wavelet-based maximum likelihood estimator introduced by (Ramsey, 1999), (Whitcher, 2000) and (Percival and Walden, 2000) and different modifications of it have been made popular by (Dacorogna et al., 2001) and (Nason, 2008).

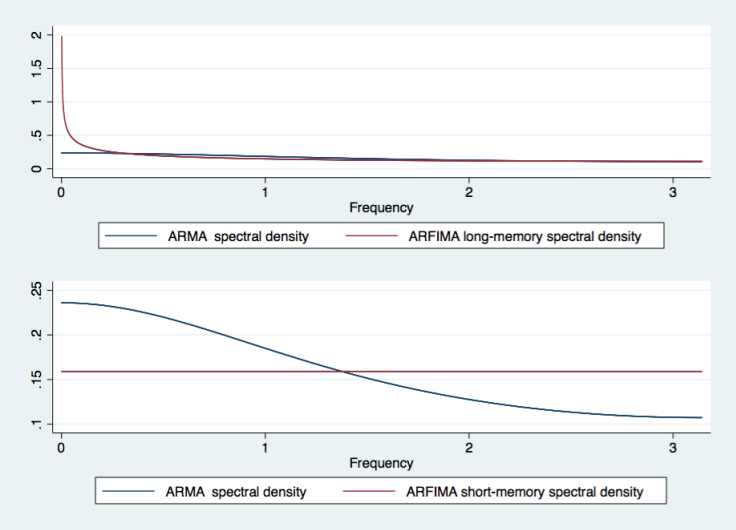

Figure 3.4: Spectral densities implied by ARFIMA parameters. The top graph contains a plot of the spectral density implied by the long-run ARFIMA parameter estimates and the ARMA(1,0) parameters. The two models imply different spectral densities for frequencies close to zero when $d > 0$. The spectral density implied by the ARFIMA estimates diverges to infinity, whereas, the spectral density implied by ARMA estimates remains finite. The bottom panel contains a graph for the short-run parameter estimates which remain finite at frequency 0. The bottom graph contains a plot of the spectral densities implied by the ARMA parameter estimates and by the short-run ARFIMA parameter estimates, which here only consists of ϵ_t.

By looking at the spectrum of the output of Discrete Wavelet Transform decompositions, and formulating the likelihood in terms of the DWT the difficulty of computing the exact likelihood is solved. We borrow from (Whitcher, 2000):

Assume, x_t is a zero mean realization from a fractional difference process with length $N = 2^j$ with spectrum given by $S_X(f) = \frac{\sigma_\varepsilon^2}{|2\sin(\pi\nu)|^{2d}}$ and covariance matrix given by $\sum_X$, The likelihood can be written as:

$$L(d, \sigma_\varepsilon^2|\mathbf{x}) = (2\pi)^{\frac{-N}{2}} |\sum|^{-1/2} \exp\left[-\frac{1}{2}\mathbf{X}^T \sum^{-1} \mathbf{X}\right] \tag{3.51}$$

And suppose that the quantity $|\sum_X|$ is the determinant of $\sum_X$. The maximum likelihood estimates of the parameters d and σ_ε^2 are those quantities which maximize equation 3.51.

The difficulties in computing the exact MLEs are avoided by using the approximate decorrelation of the DWT. If the DWT is applied on fractional difference process, then: The dependence of the likelihood function on d and σ_ε^2 is only through $\sum_X$ alone. The above likelihood function for any particular parameters values can in principle be evaluated. There are however, two practical problems with exact MLEs (Beran (1994), section 5.4). First, their determination can be very time consuming to calculate because $L(d, \sigma_\varepsilon^2|\mathbf{x})$ is computationally expensive to evaluate, even for a moderate N. Second, there can be potential numerical instabilities in computing the likelihood function when d is close to 1/2.

Consider an approximate maximum likelihood scheme that exploits the fact that $\sum_X$ can be approximated by $\widehat{\sum} = \mathbf{W}^T \digamma_N \mathbf{W}$, that is:

$$\sum_X \approx \widehat{\sum_X} = \mathbf{W}^T \digamma_N \mathbf{W} \tag{3.52}$$

where W is the orthonormal matrix defining the DWT (equation 2.49) and $\digamma_N$ is a diagonal matrix containing the variances of DWT coefficients computed from Fractional Difference processes:

$$\digamma_N = diag(N/2\underbrace{S_1, ..., S_1}, N/4\underbrace{S_2, ..., S_2}, ..., N/2^j\underbrace{S_j, ..., S_j}, ..., S_J, S_{J+1})$$

where $S_j = var(W_{j,t})$ for $j = 1, ..., J$. It is very convenient to show how S_j exactly depends on d and σ_ε^2:

$$S_j = \sigma_\varepsilon^2 S_J^{'}(d), where S_j = 2^{j+1}\int_{1/2^{j+1}}^{1/2^j} S_X(\nu)d\nu = 2^{j+1}\int_{1/2^{j+1}}^{1/2^j} \frac{\sigma_\varepsilon^2}{|2\sin(\pi\nu)|^{2d}}d\nu \quad (3.53)$$

The approximate likelihood function is now

$$\widehat{L}(d, \sigma_\varepsilon^2|\mathbf{x}) = (2\pi)^{\frac{-N}{2}}|\widehat{\sum_X}|^{-1/2}\exp\left[-\tfrac{1}{2}\mathbf{X}^T\widehat{\sum_X^{-1}}\mathbf{X}\right]$$

Hence, the task is now to find the values of d and σ_ε^2 that minimize the log-likelihood function:

$$\widehat{l}(d, \sigma_\varepsilon^2|\mathbf{x}) = -2(\log(\widehat{L}(d, \sigma_\varepsilon^2|\mathbf{x})) - N\log(2\pi)$$

$$= \log(|\widehat{\sum_X}|) + \mathbf{X}^T\widehat{\sum_X^{-1}}\mathbf{X} \quad (3.54)$$

(Whitcher, 2000) finds that equation 3.51 can be re-written as:

$$\widehat{l}(d, \sigma_\varepsilon^2|\mathbf{X}) = N\log(\sigma_\varepsilon^2) + \log(S_{J+1}^{'}(d)) + \sum_{j=1}^{J} N_j \log(S_J^{'}(d)) + \quad (3.55)$$

$$\tfrac{1}{\sigma_\varepsilon^2}(\tfrac{V_{J,0}^2}{S_{J+1}^{'}(d)} + \textstyle\sum_{j=1}^{J}\tfrac{1}{S_J^{'}(d)}\sum_{t=0}^{N_{j-1}} W_{j,t}^2)$$

We can obtain an expression for the approximate MLE $\widehat{\sigma}_\varepsilon^2$ of σ_ε^2 by differentiating the right hand side of equation 3.55 with respect to σ_ε^2 and setting the resulting expression to zero. Solving for σ_ε^2 yields the estimator $\widehat{\sigma}_\varepsilon^2$, which can be regarded as a function of d given by:

$$\widehat{\sigma}_\varepsilon^2(d) = \frac{1}{N}(\frac{V_{J,0}^2}{S_{J+1}^{'}(d)} + \textstyle\sum_{j=1}^{J}\frac{1}{S_J^{'}(d)}\sum_{t=0}^{N_{j-1}} W_{j,t}^2) \quad (3.56)$$

Using the above expression, the parameter σ_ε^2 can now be eliminated from equation 3.55. The reduced log likelihood:

$$\widehat{l}(d|\mathbf{X}) = N\log(\widehat{\sigma}_\varepsilon^2) + \log(S_{J+1}^{'}(d)) + \sum_{j=1}^{J} N_j \log(S_J^{'}(d)) \quad (3.57)$$

d		MODWT Haar	MODWT LA(8) (WMLE)
0.2	mean	0.186	0.192
0.2	SD	0.065	0.068
0.4	mean	0.346	0.384
0.4	SD	0.051	0.064

Table 3.4: Wavelet-based maximum likelihood estimator of d for $N = 2^7$.

Once equation 3.56 is subsitted into equation 3.57, the reduced likelihood depends on just the single parameter d and the MLE is obtained via $\hat{d} = arg[max\hat{l}(d|X)]$. Quite often, one is interested in long memory parameters which are strictly positive. Luckily wavelet coefficients only enter $\hat{l}(d|X)$ through quadratic forms, such as $W_j^T W_j$ and can be interpreted as estimates of the true Spectral Density Power (SDF) over the interval B_j. These estimates are then compared with the true SDF over B_j.

The process explained above is similar to the Whittle likelihood. In Whittle likelihood the true SDF is compared with an estimate of the SDF (using Fourier methods) at different frequencies. Now let $\hat{v_j^2} = W_j^T W_j/N$ denote the biased DWT-based estimator of the wavelet variance and rewrite equation 3.56, then:

$$\hat{\sigma}_\varepsilon^2 = \frac{\hat{V}_{J+1,}^2}{S'_{J+1}(d)} + \sum\nolimits_{j=1}^{J} \frac{1}{S'_J(d)} \hat{v}_j^2 \tag{3.58}$$

where $\hat{V}_{J+1} = V_J^T V_J/N$. Instead of using DWT, one can also use MODWT. (Percival and Walden, 2000) prove that using MODWT leads to a more asymptotically efficient estimator. Therefore for the estimates of this chapter the following estimator is used:

$$\tilde{\sigma}_\varepsilon^2 = \frac{\tilde{V}_{J+1,}^2}{S'_{J+1}(d)} + \sum\nolimits_{j=1}^{J} \frac{1}{S'_J(d)} \tilde{v}_j^2 \tag{3.59}$$

The simulation results provided in table 3.4 summarises the results for estimating d using MODWT, Haar wavelet and Least Asymmetric wavelets of order 8 have been considered. It is clear that MODWT with LA(8) which is called WMLE gives closer estimates to the true values of d. Fractional Difference processes have been generated using (Fraley et al., 2006). The results in table 3.4 are in the same line with the simulation results of (Whitcher, 2000) and (Percival and Walden, 2000).

Simulation of ARFIMA models

Recall that an ARFIMA(0,d,0) is presented as:

$$(1-L)^d x_t = \epsilon_t \tag{3.60}$$

Using binomial series expansion:

$$(1-L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} \tag{3.61}$$

where $\Gamma(*)$ is the gamma function. The infinite moving average representation of x_t can be denoted as:

$$x_t = (1-L)^{-d}\epsilon_t = \delta(L)\epsilon_t = \sum_{j=0}^{\infty} \delta_j \epsilon_{t-j} = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} \epsilon_{t-j} \tag{3.62}$$

Which can be used to simulate an ARFIMA(0,d,0) process. They can be also simulated using their spectral density function. Another way to simulate an ARFIMA (0,d,0) in the time domain which was first mentioned in (Granger and Joyeux, 1980) is to simulate sufficient (j) AR(1) processes with coefficients α_j for the AR(1) having a beta distribution in first step and in the second step calculate the cross-aggregate of all these AR(1) processes. To simulate an ARFIMA(p,d,q), one can use the simulated ARFIMA(0,d,0) series as the innovations for the simulation of an ARMA(p,q) process.

3.5 Other Empirical Properties of Daily Returns

The standard deviation of returns completely dominates the mean of returns at horizons such as daily. It is not possible to statistically reject a zero mean return. The daily S&P500 returns have a daily mean of 0.01791% and a daily standard deviation of 0.63272%. The mean of the returns seems to be much smaller than the standard deviation of returns. This result has important implications for the approach of modeling and measuring variation as the mean of daily returns is often replaced with zero.

The unconditional distribution of daily returns does not follow the normal distribution. Figure (3.5) shows a histogram of the daily S&P500 return data with the normal

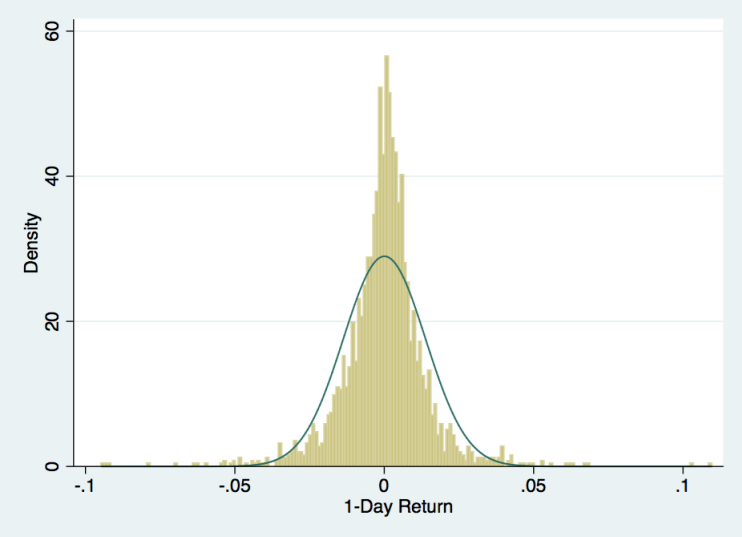

Figure 3.5: The daily S&P500 returns from 1/1/2001 through 31/12/2010 are used to construct a histogram shown in golden bars. A normal distribution with the same mean and standard deviation is shown in the blue line.

	Mean	Standard Deviation	Skewness	Kurtosis
S&P500 1-Day return	0.0000559	0.0137714	-0.1214442	11.1451
S&P500 5-day return	0.0002806	0.025602	-0.8310671	7.398657
S&P500 10-Day Return	0.0005613	0.0377041	-0.7963029	6.621245

Table 3.5: Full summary statistics of S&P500 from 1/1/2001 through 31/12/2010 at 1-Day, 5-Day and 10-Day horizons..

distribution superimposed. The histogram shows that the distribution of daily returns is more peaked around zero than the normal distribution and as expected the kurtosis is 3.765362 (see table (3.5)). Extreme returns are also more common in daily returns than in the normal distribution, a feature which is often known as "fat tails". Fat tails mean a higher probability of large losses (and gains) than the normal distribution would suggest. The skewness is negative which resembles the fact that there are very large drops but not equally large increases. The return distribution is negatively skewed and it is asymmetric. Foreign exchange markets do not show much evidence of skewness as much as stock markets do.

The means and standard deviations increase as horizon increases. On the other hand, excess kurtosis decreases. The distributions of returns change as horizon changes, perhaps at longer horizons they look more like the normal distribution.

One of the most critical decisions in the risk management industry is the importance and the difficulties in choosing the suitable time scale. (Lo and Mueller, 2010) consider the situation in which the timescales of the strategy (model) do not match the time scales of the observations (which is sometimes called a timescale mismatch). There is no "golden" timescale and the appropriate choice for the time horizon depends on the application. Day traders calculate the profit and loss daily and therefore it makes sense to calculate the risk over a time horizon of one trading day. For an investment portfolio held by a pension fund, a time horizon of one month may be chosen as the portfolio is traded less actively, some of the assets in the portfolio are less liquid and also the performance of pension fund portfolios is often monitored monthly[1]. There are two familiar ways to deal with calculating the risk at different horizons and then translating it into the desired horizon.

The question "*What is the relevant horizon for risk management ?*" has no obvious answer. There is no "natural" or "received" horizon in risk management industry [Christoffersen et al. (1998) Diebold et al. (1998) Christoffersen and Diebold (2000)]. One cannot talk of a "golden" risk horizon! Horizons can vary from very small intervals (nanoseconds, milliseconds, 5 minutes) to longer ones (1-day, 5-day, 15-days, weeks, months, quarters, or even years) depending on the motivation of trading, re-balancing the portfolio, or the asset class, industry (banking vs. insurance), position in the firm (trading desk or CFO), ... , etc.

The most common and practical rule in the risk management industry to calculate the risk at shorter horizons and then modify it into longer horizons is by scaling (Lo and Mueller, 2010). For instance, if standard deviation is calculated at a 1-hour frequency, it is transformed to a 30-day risk measure by multiplying the 1-hour standard deviation by $\sqrt{720}$ (there are 720 hours in 30 days). Sometimes, a single measure of volatility is required. For example, if R_t is daily return, then assuming financial markets are active for 252 days per year, the standard measure of annual volatility is given by $s.d.(R_t) \times \sqrt{252}$. However, this type of scaling is only valid if the underlying data is identically and independently distributed, which is not the case for financial time series at all and has been widely explored in the literature [Peter (1973), Mantegna and Stanley (1995), Diebold et al. (1998), Gencay et al. (2001b), Dacorogna et al. (2001)].

[1] 30-day horizons for interest rate instruments, 7 to 10 days for equity and foreign exchange are the timescales that are very popular.

An alternative way to modify the risk to a desired horizon is via temporal aggregation.

3.6 Aggregation

In general, in economics the ways variables aggregate could take three different forms. These are i) aggregation for a stock variable through time or systematic sampling, ii) aggregation for a flow variable or temporal aggregation, and iii) cross-section aggregation or contemporaneous,.

Systematic (skip) sampling (temporal aggregation through interval sampling), is the aggregation in which the aggregated variable is observed every h periods from the high frequency variable. For example, weekly observations may be obtained by sampling every 5 periods of daily observations. In the diagram below, if x_i is a stock variable that has a value at every i periods, z_i is a new variable observed at a different frequency:

$$\begin{array}{cc} \mathrm{x}_0 & z_0 \\ \mathrm{x}_1 & . \\ \mathrm{x}_2 & . \\ \mathrm{x}_3 & . \\ \mathrm{x}_4 & . \\ \mathrm{x}_5 & z_1 \\ \mathrm{x}_6 & . \\ \mathrm{x}_7 & . \\ \mathrm{x}_8 & . \\ \mathrm{x}_9 & . \\ \mathrm{x}_{10} & z_2 \end{array}$$

Temporal aggregation, refers to aggregation in which the aggregated variable is a flow variable and by summing the variable, at every h period, a new variable at a different frequency is formed. In the diagram below, if x_i is a flow variable, then y_i is a temporally aggregated variable:

$$\underbrace{x_0, x_1, x_2, x_3, x_4, x_5}_{y_1=\sum x_i}, \underbrace{x_6, x_7, x_8, x_9, x_{10}}_{y_2=\sum x_i}$$

Contemporaneous aggregation, is the aggregation in which individual series are aggregated and form a new time series. For example, to model the risk of a portfolio the risk models of each assets are cross aggregated.

The literature on aggregation deals with all three types of aggregation, while this chapter is only concerned about the first two types of aggregation. Aggregation has been the subject of econometric literature for at least 4 decades since (Amemiya and Wu, 1972) where they show that, if the original variable is generated by an AR model of order p, the aggregate variable follows an AR model of order p with MA residuals structure. (Granger and Joyeux, 1980) shows that under certain conditions a linear combination of an infinite set of AR(1) processes is a fractionally integrated process which exhibit long memory. (Beran, 1994) uses this technique to simulate long-memory processes. (Drost and Nijman, 1993) derive the order of conditions for temporally aggregated univariate GARCH models. (Chambers, 1998) considers the cross-section and temporal aggregation of macroeconomic time series in which he compares the spectral densities of aggregated series with that of individuals. (Granger, 1980) looks into contemporaneous aggregations of more general long-memory and short-memory processes. (Marcellino, 1999) reviews several consequences of temporal aggregation for the class of ARMA processes. (Zaffaroni, 2007a) and (Zaffaroni, 2007b) study the effect of Contemporaneous aggregation of GARCH processes with both cases of dependent and independent individuals. (Silvestrini and Veredas, 2008) provide a complete survey of temporal aggregation of univariate and multivariate series however, they only focus on ARMA and GARCH models and they do not consider ARFIMA models. They do not study how the autocorrelation functions change over temporal aggregation and they limit the study to stationary short-memory processes, but still the survey sums up the literature on temporal aggregation of ARMA and GARCH models. Most of the studies (either in finance or macroeconometrics) focus on cross-aggregation of short-memory processes rather than on temporal aggregation, with the exception of (Chambers, 1998), (HWANG, 2000), (Silvestrini and Veredas, 2008) and a series of studies followed up from (Souza and Smith, 2002).

There are however few studies that compare the properties of long-memory processes at different frequencies. (HWANG, 2000) follows the steps of (Chambers, 1998) and find that in short lags the autocorrelation functions are affected as an outcome

of aggregation, whereas, in large lags the autocorrelation functions remian unaffected. (Andersen et al., 2001b) compare the properties of volatility and autocorrelation functions of Deutschemark and Yen returns against the dollar for 10 years and find a series of scaling laws across different requencies. (Ohanissian et al., 2008) propose a test statistic to distinguish between the true long-memory processes and spurious long memory processes comparing the statistical properties of long-memory time series before and after sampling. (Ahmad and Paya, 2013) study the effects of time averaging and interval sampling for the case of random walk.

A natural question to ask is since there are different risk horizons used in the risk industry, and since scaling is not the appropriate method to use, what will happen if simply one chooses the corresponding sampling frequency? If the original variable shows signs of long-memory with $d = d_1$, will the time series sampled at a different frequency also be the case of long-memory with $d = d_1$ or will it be a different d?

3.6.1 Temporal aggregation or skip sampling?

Recall from equation 4.3 that $R_t = ln(P_t) - ln(P_{t-1} = ln(\frac{P_t}{P_{t-1}})$ is the log return for the $t-th$ day if P_t is the closing price at day t and P_{t-1} is the closing price at dat $t-1$. The log return series for a 5-day non-overlapping holding period can be constructed by:

$$\mathrm{R}_{5t} = ln(\frac{P_t}{P_{t-5}})$$

$$\mathrm{R}_{5t} = ln[(\frac{p_t}{p_{t-1}})(\frac{p_t-1}{p_{t-2}})(\frac{p_t-2}{p_{t-3}})(\frac{p_t-3}{p_{t-4}})(\frac{p_t-4}{p_{t-5}})]$$

$$\mathrm{R}_{5t} = ln(\frac{p_t}{p_{t-1}}) + ln(\frac{p_t-1}{p_{t-2}}) + ln(\frac{p_t-2}{p_{t-3}}) + ln(\frac{p_t-3}{p_{t-4}}) + ln(\frac{p_t-4}{p_{t-5}})$$

$$\mathrm{R}_{5t} = \sum_{t=1}^{5} R_t$$

Which shows that in fact when calculating the return at different horizons, it is by temporal aggregation that we find the return series at different frequencies.

3.6.2 The Effects of Temporal Aggregation on the Properties of ARFIMA (0,d,0) Models

In risk management there are studies that use different frequencies to calculate the risk, from monthly observations such as (Jacobsen and Dannenburg, 2003) to ultra high frequency data in (Martens, 2002).

The effects of skip sampling on the properties of ARFIMA models (a downward bias on $\hat{d}$) have been discussed in details for example in (Souza and Smith, 2002) and are not discussed here, but the effects of temporal aggregation is dealt with in much details. Equation (3.37) plays an important role for long-memory processes as it shows that the rate at which the autocorrelation functions decrease is slower than short-memory processes. In this section we calculate the autocorrelation functions of a temporally aggregated ARFIMA(0,d,0) and compare it with that of not aggregated series. Then the 5-day and the 10-day S&P500 returns are calculated and tested against the results of temporal aggregation.

We assume that the sampling interval is longer than the true interval. More precisely we assume that the original time series shown as x_t happens at every unit of time and the time series y_t is observed at every h unit of time. For example in the case of 5-day returns, $h = 5$, the true variable happens at every day ($h = 1$) unit of time but we observe the variable every 5 days, so $h = 5$. The temporally aggregated variable y_t is constructed as:

$$y_{ht} = y_T = x_t + x_{t-1} + x_{t-2} + \ldots + x_{t-(h-1)=\sum_{i=o}^{h-1} x_{t-i}} \quad (3.63)$$

where h is the numbers aggregated together. Here t represents units of fundamental time and $T = ht$ units of aggregate time. For example, if x_t is is the daily returns, and $h = 5$, then one unit of T goes through 5 units of t time. The τ lag of y_T is given be:

$$Y_{T-\tau} = x_{t-h\tau} + x_{t-(h\tau+1)} + \ldots + x_{(t-(h+1)\tau-1)} = \sum_{i=h\tau}^{(h+1)\tau-1} x_{t-i} \quad (3.64)$$

For the 5-day return:

$$\underbrace{x_t, x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}}_{y_T}, \underbrace{x_{t-5}, x_{t-6}, x_{t-7}, x_{t-8}, x_{t-9}}_{y_{T-1}}$$

Hence, $L.y_T = y_{T-1} = x_{t-5} + x_{t-6} + x_{t-7} + x_{t-8} + x_{t-9}$. To give an intuition behind the idea of finding autocorrelation functions of an aggregated process, we begin with the case $d = 0$ (short-memory processes) and summarise the results of (Tiao, 1972). The autocorrelation between y_T and y_{T-1} is:

$$\gamma_1^y = \frac{cov(y_T, y_{T-1})}{var(y_T)} = \frac{E[(x_t + \ldots + x_{t-(h-1)})(x_{t-h} + \ldots + x_{t-(2h-1)})]}{E(x_t + \ldots + x_{t-(h-1)})^2} \tag{3.65}$$

The terms in the numerator and the denominator both consist of products of lagged x_t. However, the numerator involves observations of x that are on average further distant from each other compared to the denominator. For γ_τ as τ increases, this distance increases. Since $d = 0$, the expectation of the terms in the numerator converges to zero faster, on average, than those in the denominator, bringing the autocorrelation to zero as τ increases. Therefore the autocorrelations of the temporally aggregated short-memory processes approach zero as h increases as the numerator converges to zero faster.

If $d = 1$, that is x_t is the first difference stationary, because temporal aggregation is a linear filter it preserves the number of unit roots and hence y_T is also a first difference stationary. If $z = (1 - L)y$, then the autocorrelation between z_T and $z - T - 2$ is:

$$\gamma_1^z = \frac{cov(z_T, z_{T-2})}{var(z_T)} = \frac{E[(y_T - y_{T-1})(y_{T-2} - y_{T-3})]}{E(y_T - y_{T-1})^2} \tag{3.66}$$

Using a similar reasoning leads to a fast coverging to zero autocorrelation functions which further summs up the matter for aggregation when $d = 0$ and $d = 1$. What we are interested in what changes (if any) when a long-memory process is aggregated. Recall from equation 3.62 that an ARFIMA process can be denoted as:

$$x_t = (1 - L)^{-d}\epsilon_t = \sum_{j=0}^{\infty} \frac{\Gamma(j + d)}{\Gamma(j + 1)\Gamma(d)}\epsilon_{t-j} \tag{3.67}$$

And if an aggregated process y_{ht} is constructed through $y_T = x_t + x_{t-1} + x_{t-2} + \ldots + x_{t-(h-1)}$, then:

$$
\begin{aligned}
\mathrm{y}_{ht} &= \textstyle\sum_{i=0}^{i=h-1}(1-L)^{-d}\epsilon_{t-i} \\
y_T &= (1-L)^{-d}\epsilon_t + (1-L)^{-d}\epsilon_{t-1} + (1-L)^{-d}\epsilon_{t-2} + \ldots + (1-L)^{-d}\epsilon_{t-(h-1)} \\
y_T &= (1-L)^{-d}[\epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} + \ldots + \epsilon_{t-(h-1)}] \\
y_T &= (1-L)^{-d}[1 + L + L^2 + \ldots + L^{h-1}]\epsilon_t \\
y_T &= (1-L)^{-d}[1 + L + L^2 + \ldots + L^{h-1}]\epsilon_t \times \frac{(1-L)}{(1-L)} \\
y_T &= (1-L)^{(-d-1)}[1 - L^h]\epsilon_t
\end{aligned}
$$

And hence the aggregated series y_T has the form of:

$$y_T = \sum_{j=0}^{h-1}(1-L)^{-d}\epsilon_{t-j} = (1-L)^{(-d-1)}[1-L^h]\epsilon_t \tag{3.68}$$

In ARFIMA processes the most important part of the model is the parameter "d", which shows how persistent the memory of the variable is. The most important and natural step would be to compare the persistence of the aggregated process with that of the original series and check if aggregation causes an increase or a decrease in the decay rate of the autocorrelation functions. To see whether aggregation has any effect on ARFIMA processes, the autocovariance, autocorrelation and spectral density functions of the aggregated process is compared with that of the original one, i.e., properties of x_t are compared against y_T.

The autocovariance ($\kappa^y(\tau,h)$), autocorrelation ($\gamma^y(\tau,h)$, and spectral density ($f^y(\nu,h)$) functions of y_T are given by:

$$\kappa^y(\tau,h) = \frac{\sigma^2\Gamma(1-2d)}{\Gamma(d+1)\Gamma(1-d)2(1+2d)} \times [\frac{\Gamma(1+h\tau+d-h)}{\Gamma(h\tau-d-h)} + \frac{\Gamma(1+h\tau+d+h)}{\Gamma(h\tau-d+h)} - 2\frac{\Gamma(1+h\tau+d)}{\Gamma(h\tau-d)}] \tag{3.69}$$

$$\gamma^y(\tau,h) = \frac{[\frac{\Gamma(1+h\tau+d-h)}{\Gamma(h\tau-d-h)} + \frac{\Gamma(1+h\tau+d+h)}{\Gamma(h\tau-d+h)} - 2\frac{\Gamma(1+h\tau+d)}{\Gamma(h\tau-d)}]}{[2\frac{\Gamma(1+d-h)}{\Gamma(-d-h)} - 2\frac{\Gamma(1+d)}{\Gamma(-d)}]} \tag{3.70}$$

$$f^y(\nu,h) = \frac{\sigma^2}{2\pi}\left(2sin(\frac{\nu}{2h})\right)^{-2(d+1)}(2sin(\frac{\nu}{2})^2 \tag{3.71}$$

Where h is a sampling interval, τ is the lag between observations, and ν is the frequency.

Proof: See the Appendix.

If h is relatively large using Stirling's approximation, equation (3.70) can be re-written as[1]:

$$\gamma^y(\tau,h) \sim \frac{(h\tau+h)^{2d+1}+(h\tau-h)^{2d+1}-2h\tau^{2d+1}}{2h^{2d+1}-2\dfrac{\Gamma(1+d)}{\Gamma(-d)}}$$

$$\gamma^y(\tau,h) \sim \frac{h^{2d+1}[(\tau+1)^{2d+1}+(\tau-1)^{2d+1}-2\tau^{2d+1}]}{2h^{2d+1}-2\dfrac{\Gamma(1+d)}{\Gamma(-d)}} \tag{3.72}$$

Equation $\lim_{\tau\longrightarrow\infty}\gamma^y(\tau,h)$ when $\tau\longrightarrow\infty$, is a fraction that has a numerator of a $\infty+\infty-\infty$ case for which the common factor with the greatest exponent should be removed from the polynomial. Using, $(x+y)^n=\sum_{k=0}^{n}\binom{n}{k}x^{n-k}y^k$ and diving up the numerator:

$$(\tau+1)^{2d+1}=$$
$$\tbinom{2d+1}{0}\tau^{2d+1}1^0+\tbinom{2d+1}{1}\tau^{2d}1^1+\tbinom{2d+1}{2}\tau^{2d-1}1^2+\tbinom{2d+1}{3}\tau^{2d-2}1^3+\ldots+\tbinom{2d+1}{2d+1}\tau^0 1^{2d+1}=$$
$$\tau^{2d+1}\left[1+(2d+1)\frac{1}{\tau}+d(2d+1)\frac{1}{\tau^2}+\frac{(2d+1)d(2d-1)}{3}\frac{1}{\tau^3}+\ldots+\frac{1}{\tau^{2d+1}}1^{2d+1}\right]$$

And

$$(\tau-1)^{2d+1}=\tbinom{2d+1}{0}\tau^{2d+1}(-1)^0+\tbinom{2d+1}{1}\tau^{2d}(-1)^1+\tbinom{2d+1}{2}\tau^{2d-1}(-1)^2+$$
$$\tbinom{2d+1}{3}\tau^{2d-2}(-1)^3+\ldots+\tbinom{2d+1}{2d+1}\tau^0(-1)^{2d+1}=$$
$$\tau^{2d+1}\left[1-(2d+1)\frac{1}{\tau}+d(2d+1)\frac{1}{\tau^2}-\frac{(2d+1)d(2d-1)}{3}\frac{1}{\tau^3}+\ldots+\frac{1}{\tau^{2d+1}}(-1)^{2d+1}\right]$$

[1]Stirling's approximation is $\Gamma(z)\sim z^{z-\frac{1}{2}}e^{-z}\sqrt{2\pi}$ which is used for $\frac{\Gamma(z+a)}{\Gamma(z+b)}\sim z^{a-b}$ for large z.

And so $(\tau+1)^{2d+1}+(\tau-1)^{2d+1}-2\tau^{2d+1}$ can be written as:

$$(\tau+1)^{2d+1}+(\tau-1)^{2d+1}-2\tau^{2d+1}=$$
$$\tau^{2d+1}\left[1+(2d+1)\frac{1}{\tau}+d(2d+1)\frac{1}{\tau^2}+\frac{(2d+1)d(2d-1)}{3}\frac{1}{\tau^3}\right.$$
$$\left.+\ldots+\tau^0 1^{2d+1}+1-(2d+1)\frac{1}{\tau}+d(2d+1)\frac{1}{\tau^2}-\frac{(2d+1)d(2d-1)}{3}\frac{1}{\tau^3}+\ldots+\tau^0(-1)^{2d+1}-2\right]$$
$$=\tau^{2d+1}\left[2d(2d+1)\frac{1}{\tau^2}+\ldots\right]$$
$$=2\mathrm{d}(2\mathrm{d}+1)\tau^{2d-1}$$

And so:

$$\lim_{\tau\longrightarrow\infty}\frac{h^{2d+1}[(\tau+1)^{2d+1}+(\tau-1)^{2d+1}-2\tau^{2d+1}]}{2h^{2d+1}-2\dfrac{\Gamma(1+d)}{\Gamma(-d)}}$$
$$=\tau^{2d-1}\times\frac{h^{2d+1}\left[2d(2d+1)\right]}{h^{2d+1}-2\dfrac{\Gamma(1+d)}{\Gamma(-d)}}$$
$$=\tau^{2d-1}\times\frac{h^{2d+1}\left[d(2d+1)\right]}{h^{2d+1}-\dfrac{\Gamma(1+d)}{\Gamma(-d)}}$$

If $\tau\longrightarrow\infty$, $\lim\gamma^y(\tau,h)$ will be:

$$\lim\gamma^y(\tau,h)=\tau^{2d-1}\times\frac{h^{2d+1}d(1+2d)}{h^{2d+1}-\dfrac{\Gamma(1+d)}{\Gamma(-d)}} \tag{3.73}$$

For when $\nu\longrightarrow 0$, $\lim f^y(\nu,h)$ is:

$$\lim_{\nu\longrightarrow 0}\frac{\sigma^2}{2\pi}\left(2sin(\frac{\nu}{2h})\right)^{-2(d+1)}(2sin(\frac{\nu}{2}))^2=$$

If $x \longrightarrow 0$, then $sin(x) \sim x$. If $\nu \longrightarrow 0$:

$$\lim_{\nu \longrightarrow 0} \frac{\sigma^2}{2\pi} \left(2sin(\frac{\nu}{2h})\right)^{-2(d+1)} \left(2sin(\frac{\nu}{2}\right)^2 = \frac{\sigma^2}{2\pi} \left(2\frac{\nu}{2h}\right)^{-2d-2} \times \left(2\frac{\nu}{2}\right) = \frac{\sigma^2}{2\pi}(h)^{2d+2}\nu^{-2d-2}\nu^2 = \frac{\sigma^2}{2\pi}(h)^{2d+2}\nu^{-2d}$$

$$\lim f^y(\nu, h) = \frac{\sigma^2}{2\pi} h^{2d+2}\nu^{-2d} \tag{3.74}$$

Equations (3.73) and (3.74) in comparison with equations (3.37) and (3.38) show that the decay rates of the aggregated ARFIMA(0,d,0) and ARFIMA(0,d,0) are the same, τ^{2d-1} (similar to the results in (Chambers, 1998), and inline with (Souza and Smith, 2004), (Souza, 2005) and (Hassler, 2011)) but the level at which they start from are different. Since:

$$F = \frac{F_1}{F_2} = \frac{\tau^{2d-1} \times \dfrac{h^{2d+1}d(1+2d)}{h^{2d+1} - \dfrac{\Gamma(1+d)}{\Gamma(-d)}}}{\dfrac{\tau^{2d-1} \times \Gamma(1-d)}{\Gamma(d)}} \tag{3.75}$$

$$= \frac{\dfrac{h^{2d+1}d(1+2d)}{h^{2d+1} - \dfrac{\Gamma(1+d)}{\Gamma(-d)}}}{\dfrac{\Gamma(1-d)}{\Gamma(d)}} \tag{3.76}$$

In equation (3.37), $\frac{\Gamma(1-d)}{\Gamma(d)}$ is always smaller than $\frac{h^{2d+1}d(1+2d)}{h^{2d+1} - \frac{\Gamma(1+d)}{\Gamma(-d)}}$ in equation (3.73) for positive d. Table (3.6) summarises the values of F_1 and F_2 for different (positive) values of d and h:

Even though the limiting autocorrelation functions have the same decay rate, the level of autocorrelation function is biased upwards, hence making the observed (at different sampling frequency) aggregated ARFIMA(0,d,0) processes more persistent in comparison to the true underlying process. There are two immediate outcomes from table (3.6) and comparing equations (3.37) and (3.73). First, F_1 is greater F_2 for all

F1	F2	F1	F2
d=0.4, h=5		d=0.4, h=10	
0.717466	0.601479	0.717105	0.601479
d=0.3, h=5		d=0.3, h=10	
0.477092	0.348327	0.477587	0.348327
d=0.2, h=5		d=0.2, h=10	
0.277908	0.182454	0.278740	0.182454
d=0.1, h=5		d=0.1, h=10	
0.119208	0.072424	0.119654	0.072424

Table 3.6: Different values of F_1 and F_2 depending on different values of d and h from equations (3.37) and (3.73).

values of d, i.e., the aggregated process is biased upwards compared to the true process. Second, as h increases, F_1 also increases, i.e., as sampling frequency increases the degree of persistence in the aggregated process seem to have increased.

3.6.2.1 Monte Carlo Replications

Four different series (all following an ARFIMA(0,d,0) process) each with a sample size of $N = 2000$ are generated. Since the autocorrelation functions of the aggregated processes depend on the values of d and h, the 4 simulated series each correspond to a different value of d (d=0.4, 0.3, 0.2, 0.1). The original series are then aggregated via $y_{ht} = \sum_{i=o}^{h-1} x_{t-i}$ for sampling frequencies of h=1, 5, 10 and 20 with sample sizes of 2000, 400, 200 and 100 respectively. For each combination of d and h, $\hat{d}$ has been estimated using the following three methods (1) GPH for $m = 0.5$, (2) WMLE and (3) Whittle[1]. This procedure was carried out 2000 times and $\hat{d}$ was calculated in each repetition. Tables (3.7) and (3.8) summarises the averages and the standard deviations of $\hat{d}$ in 2000 repetition.

As can be seen from tables (3.7) and (3.8), the aggregated ARFIMA(0,d,0) processes show the upward bias as the effect of aggregation. The value of d of the aggregated ARFIMA(0,d,0) process is greater than the value of d of the actual ARFIMA(0,d,0) process for all values of d. Out of the three estimators, GPH seems to be the one that gives estimates closer to the true value of d, however it has the highest standard deviation among the Whittle and WMLE estimators. The Whittle estimator and the

[1]The method introduced by (Sowell, 1992a) is very computationally expensive for a sample size of 1000, and has not been used here.

True d	Sampling Interval				Sampling Interval			
	GPH				WMLE			
	1	5	10	20	1	5	10	20
0.4	**0.4017**	**0.4038**	**0.4029**	**0.4125**	0.4021	0.4216	0.4641	0.4578
0.3	**0.2985**	**0.2962**	**0.2937**	**0.2964**	0.2971	0.3375	0.3767	0.3698
0.2	**0.1983**	**0.2020**	**0.2076**	**0.2052**	0.2019	0.2494	0.2656	0.2646
0.1	**0.1009**	**0.1267**	**0.1307**	**0.1262**	0.0880	0.1023	0.1466	0.1361

True d	Sampling Interval			
	Whittle			
	1	5	10	20
0.4	0.4068	0.4398	0.4663	0.4515
0.3	0.3033	0.3347	0.3890	0.3757
0.2	0.2068	0.2213	0.2398	0.2554
0.1	0.1017	0.1145	0.1290	0.1461

Table 3.7: The average value of $\hat{d}$ by three estimation methods (N=2000, 400, 200 and 100). The results are based on 2000 replications.

WMLE have much lower standard deviations (as discussed in (Gonzaga and Hauser, 2011)) and although they have given good estimations but they are still worse than GPH. It is very interesting that WMLE has standard deviations between the Whittle estimator and the GPH estimator, and even though WMLE estimator has given estimations worse than the GPH estimator, it is still doing better than the Whittle estimator.

Recall from equation (3.43) that GPH only concentrates on low frequencies (by choosing the appropriate $g(N)$) and since d is obtained by the spectral density near zero frequencies, perhaps GPH is a more suitable method for estimating the long memory parameter d in an aggregated ARFIMA process (as temporal aggregation does not affect the decay rate of the true long memory process). The estimation in GPH is very close to the true values of d, however the standard deviation is very high when the sampling interval is 20. Since the number of periodograms used in GPH estimation is only $N^{0.5}$, this was expected.

Based on the three estimators, how close the estimations are to the true values of d (bias) depends on the sampling intervals. The estimations become more distant from the true values of d as h increases from 1 to 10. However, as h reaches 20, in several occasions (for instance: GPH, 0.4 and WMLE, 0.4), the estimations either show little increase or even a decrease. It seems there is no relationship between the sampling

True d	Sampling Interval				Sampling Interval			
	GPH				WMLE			
	1	5	10	20	1	5	10	20
0.4	0.089	0.278	0.314	0.412	0.052	0.165	0.235	0.349
0.3	0.089	0.277	0.313	0.414	0.053	0.164	0.236	0.348
0.2	0.088	0.276	0.314	0.413	0.053	0.165	0.236	0.351
0.1	0.089	0.277	0.314	0.415	0.050	0.165	0.236	0.361
True d	Sampling Interval							
	Whittle							
	1	5	10	20				
0.4	**0.035**	**0.098**	**0.214**	**0.314**				
0.3	**0.038**	**0.096**	**0.212**	**0.316**				
0.2	**0.049**	**0.112**	**0.213**	**0.349**				
0.1	**0.039**	**0.099**	**0.214**	**0.317**				

Table 3.8: The Standard Deviation of $\hat{d}$ by three estimation methods (N=2000, 400, 200 and 100). The results are based on 2000 replications.

interval and $\hat{d}$. Recall from equation (3.73) that:

$$\lim \gamma^y(\tau, h) = \tau^{2d-1} \times \frac{h^{2d+1} d(1+2d)}{h^{2d+1} - \dfrac{\Gamma(1+d)}{\Gamma(-d)}}$$

As $h \to \infty$ then (a $\frac{\infty}{\infty}$ case):

$$\gamma^y(\tau, h) = d(1+2d) \times \tau^{2d-1} \tag{3.77}$$

Which then does not depend on h and only depends on the values of d and τ. Since for long memory stationary time series $0 < d < 0.5$, $0 < 2d < 1$, and hence $-1 < 2d - 1 < 0$. If $\tau \to \infty$ and $-a = 2d - 1$ then:

$$lim_{\tau \to \infty} d(1+2d) \times \tau^{2d-1} = lim_{\infty \to \infty} \frac{d(1+2d)}{\tau^a} \to 0 \tag{3.78}$$

Equation 3.78 "slowly" reaches zero as $\tau \to \infty$ since $0 < a < 1$. ARFIMA processes are models that are well suited the class of time series that have a slow decaying rate in their autocorrelation functions.

	Return	Squared Return	Absolute Return
1-Day S&P500	1.19	2.94	3.83
5-Day S&P500	2.22	2.86	2.44
10-Day S&P500	1.62	2.43	2.22

Table 3.9: The summary of the Modified R/S test statistic for the return, squared return and absolute return at 1-Day, 5-Day and 10-Day frequencies of $S\&P500$ from 1/1/2001 through 31/12/2010. The Critical Values are as follows: 90% : $[0.861, 1.747]$, 95% : $[0.809, 1.862]$, 99% : $[0.721, 2.098]$

3.6.3 S&P500 returns, 1-Day, 5-Day and 10-Day

The S&P500 daily returns are sampled at 5-day and 10-day intervals. Table 3.9 summarises the R/S test statistics for the return, squared returns and absolute returns at 1-Day, 5-Day, and 10-Day intervals. The squared and absolute returns at all sampling intervals reveal signs of long-range dependence. The daily returns and 10-Day returns do not show any signs of long-range dependence, however 5-Day returns seems to reject the H_0 of no long-range dependence.

The descriptive statistics of the data are presented in table (3.5) as well as at table (3.10) to ease the comparison. The last two columns in table (3.10) are $\hat{d}$s in ARFIMA(0,d,0) and ARFIMA(0,d,1) respectively.

n	Mean	Stan. Dev.	Skew.	Kurt.	$Q(40)^a$	$Q^2(40)^b$	$\hat{d}^c$	$\hat{d}^d$
				1-Day Return				
2513	0.000056	0.013771	-0.121444	11.1451	147.9939	5784.3449	0.1952 (0.0113)	0.4932 0.0094
				5-Day Return				
502	0.000281	0.025602	-0.831067	7.398657	102.4842	677.6419	0.2358 0.0418	0.4601 0.0049
				10-Day Return				
251	0.000561	0.037704	-0.796303	6.621245	99.2318	493.9623	0.4984 0.0022	0.4973 0.0020

Table 3.10: Full summary statistics of $S\&P500$ returns from 1/1/2001 through 31/12/2010 at three different frequencies. **a:** Ljung-Box test statistics for up to fortieth order serial correlation in returns. **b:** Ljung-Box test statistic for up to fortieth order serial correlation in squared returns. **c:** $\hat{d}$ in ARFIMA(0,d,0) for R^2, the numbers in the brackets are the standard errors. **d:** $\hat{d}$ in ARFIMA(0,d,1) for R^2, the numbers in the brackets are the standard errors.

As sampling frequency increases the mean and standard deviations of return in-

crease, while the kurtosis decreases. The 5-day returns seem to be the most negatively skewed one, while the 1-day return has a skewness closest to zero. The properties of unconditional returns are all in line with the numerous empirical evidence in the literature. The GPH estimator was used for estimating the long memory parameter d in ARFIMA(0,d,0) since it was the one giving closer estimations to the true values of d in table (3.7). The upward bias in $\hat{d}$ is obvious as the sampling interval increases which was expected. A useful route to take (suggested by (Man and Tiao, 2006)) when it comes to estimating the long memory parameter in temporally aggregated processes is (1) make sure that the short-memory parameters at the first sampling interval are estimated correctly (2) if ARFIMA(0,d,0) is the correct model for non-aggregated data but ARFIMA(0,d,0) prove to be very sensitive to the sampling intervals use $ARFIMA(0, d, d_1)$ where $d_1 = \lfloor d + 1 \rfloor$. (Man and Tiao, 2006) also found that ARFIMA(0,d,1) seems to be the most suitable and reasonable model for temporally aggregated data. Indeed this is further confirmed as in the last column of table 3.10, $\hat{d}$ s have relatively smaller standard errors. It seems that the upward bias caused by temporal aggregation can be explained by increasing the parameters in the MA part of the model. Overall, temporal aggregation does not change the long-memory characteristics of the data and if the actual data is classified as long range dependent, then there is a very good chance that observed data any sampling frequencies will also be able to reveal this characteristic of the data as long as there is enough data points left.

The effects of temporal aggregation on intraday data will be considered in the next chapter, which is in-line with the results of this chapter.

4

Volatility Modeling: Using Intra-day Observations

4.1 Introduction

With the advent of high frequency data, there have been questions raised regarding empirical findings specific to tick-by-tick data and whether conventional volatility models are capable of capturing these features. There have been several studies documenting shortcomings of GARCH models in the context of high frequency data. Andersen and Bollerslev (1998a), (Andersen et al., 2001a), (Cont, 2001), (Fleming et al., 2003), (Maheu, 2005), and (Andersen et al., 2005) emphatically raise criticisms of GARCH-type models. In particular (Andersen and Bollerslev, 1998b) and several others following them, explore the shortcomings of GARCH models when frequency of the data switches between high and low.

With the advancements in technology, high frequency data which are the building blocks of financial markets, have become increasingly available. Market prices are completely described by tick-by-tick data. Each "tick" is one unit of information that shows somewhere, someone has made a quote/transaction. High frequency data are now available due to advancement in technology and as a consequence, academics tend to use the newly available data. There are compelling reasons why one should consider using high frequency data, out of which the following reasons are those that this thesis believes are worth mentioning.

- Statistically, the higher the number of observations, the higher the degrees of freedom which will then result in more precise estimators.
- Tick-by-tick data are the building blocks of financial markets. If the aim is to analyse and explore financial time series data to the possible level of accuracy, then one should be using higher frequencies data rather than lower frequencies (daily or monthly) data.
- High frequency data make it possible to study financial markets at different time scales, from nanoseconds and minutes to years and decades which makes it possible to observe the previous latent and unobservable variables such as volatility.
- The analysis of high frequency data helps to clarify the behaviour of intraday traders which own a great share of the volume of the market. According to Reuters in 2013, high frequency trading accounts for about 40% of spot trading in currencies, up to 70% in U.S. equities and 45% in stocks globally.

All the reasons above point out to the fact that studying intraday data can be very rewarding and could possibly unravel the previous unknown relationships between several financial concepts. However widely approved empirical regularities of daily/weekly data do not hold up at data analysis of higher frequencies. One of the major differences is that the heterogeneity of market agents in high frequency data has a much higher degrees of freedom in comparison with daily/monthly data. The models which are based on the assumption of homogeneity do not work in the context of high frequency data which have been mentioned at numerous occasions for example: [Andersen and Bollerslev (1997a), Goodhart and O Hara (1997), Bollerslev and Wright (2000), Dacorogna et al. (2001), Russell et al. (2009), Lo and Mueller (2010)].

The goal of this chapter of the thesis is to explore the properties of intraday returns, to achieve a better understanding of volatility modeling using intraday data and to build up a framework for computing daily volatility, using the information in intraday returns. In doing so, the concept of realised volatility using intraday data is defined and contrasted with that of volatility using daily data. The novel part of this chapter investigates in details another feature of the data in which, intraday (higher frequencies) and daily (lower frequencies) data behave contradictory. While chapter three has discussed the effects of temporal aggregation on long memory processes at

daily and lower frequencies, chapter four argues the case for long memory processes of higher frequencies at intraday levels and provides alternative explanations for what is puzzling in the literature. Using intraday GBP/USD exchange rates for over two years, a comparison between available models for realised volatilities are made.

4.2 Data

The foreign exchange rate market is the largest financial market worldwide and it is very unique since it is 24 hours active and the 5-minute GBP/USD exchange rates have been used.

Tick-by-tick data are the building blocks of financial markets which are considered to be of ultra-high frequency. There was a time that perhaps 5-minute and 1-minute observations were called high frequency but not any more. This thesis however often calls intraday data high frequency simply because it is at a frequency higher than daily data. Consider tick-by-tick [1] observations of GBP/USD for the month of May 2010 that has a sample size of $5,436,255$. Table (4.1) shows the bid-ask spreads on 1 May 2010, within 00:00:01 (within the first second of 1 May 2010) where all the 15 rows show the changes occurring during that 1-second interval. The data is measured to the nearest millisecond (has the timestamp of millisecond)[2].

If price P_t is defined as the average of logarithmic bid and ask quotes[3]:

$$P_t = \frac{log(P_{t,bid}) + log(P_{t,ask})}{2} \tag{4.1}$$

Equation (4.1) has the advantage of behaving symmetrically when the price is inverted, i.e., \$1 expressed in GBP instead of 1 expressed in USD. Tick-by tick Price movements of GBP/USD is plotted in figure (4.7)

Using the definition of autocorrelation in equation (3.10) for the tick-by-tick GBP/USD return series, the autocorrelation functions are plotted in figure (4.2) which shows that the autocorrelation function of the returns converges to zero in a few lags. The negative autocorrelations at the very short lags in the tick-by-tick return series have been

[1] Tick-by-tick data is often known as ultra-high frequency data in the literature.

[2] Data is available at: http://www.truefx.com

[3] The definition of the price as the average of logarithmic bid and ask quotes is very similar to defining the price as the logarithm of the average, and the numerical difference between these two definitions is insignificant as explained in (Ulrich et al., 1990)

Date	Time	Bid	Ask
20100501	00:00:00.250	1.55315	1.55413
20100501	00:00:00.251	1.55322	1.55418
20100501	00:00:00.469	1.55313	1.55413
20100501	00:00:00.470	1.55313	1.55410
20100501	00:00:00.472	1.55342	1.55378
20100501	00:00:00.474	1.55310	1.55410
20100501	00:00:00.476	1.55319	1.55415
20100501	00:00:00.477	1.55335	1.55385
20100501	00:00:00.562	1.55340	1.55378
20100501	00:00:00.790	1.55308	1.55406
20100501	00:00:00.791	1.55333	1.55381
20100501	00:00:00.987	1.55340	1.55374
20100501	00:00:00.996	1.55338	1.55374
20100501	00:00:00.998	1.55306	1.55406
2010 0501	00:00:00.999	1.55314	1.55411
⋮	⋮	⋮	⋮

Table 4.1: The first 15 rows of GBP/USD Bid-Ask spread, 1 May 2010. Time is measured to the nearest millisecond.

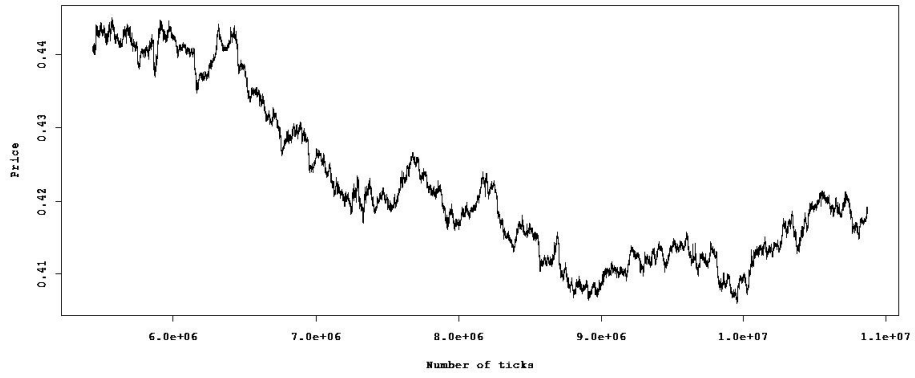

Figure 4.1: Price movements of tick by tick GBP/USD exchange rate for the whole month of May 2010.

widely observed in the literature. The negative autocorrelation at the first lag suggests a fast mean reversion of the price at the tick level which is due to the action of market makers [(Goodhart and O Hara, 1997), (Cont et al., 1997), (Cont, 2001)]. The justification behind the convergence in figure (4.2) is quite simple. If returns reveal significant correlations, then these correlations can be used to form a simple trading strategy with positive returns. However, such strategies reduce the correlations, except in very short time scales which represents the time the market takes to react to new information. This correlation time is typically several minutes and could be even shorter for foreign exchange markets that have a high volume of trading everyday. In figure (4.2), it is safe to assume that for $\tau > 68$ the correlations are zero which means it roughly takes 4 seconds for the market to incorporate new piece of information.

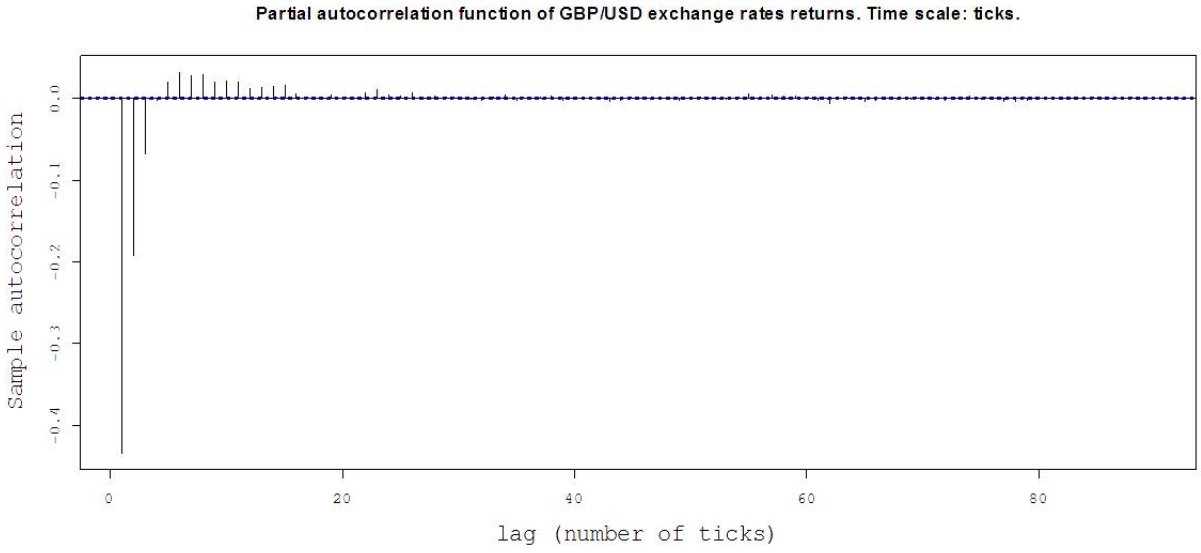

Figure 4.2: Autocorrelation Function ($\gamma(\tau)$) of tick data. The confidence intervals are the dashed blue lines.

Using all the available information at tick-by-tick level seems to lead to efficient estimators, unfortunately the returns at the very highest frequencies (tick-by-tick) can be influenced by quoting intensity. In order to minimize the effects related to the process of price formation and to avoid micro-structure noise, the literature suggests subsampling at intraday levels, i.e., using frequencies that are still within a day as long as these subsamples series are noise free. Some parts of this chapter follow the norm of using 5-minute data which is widely used in the literature [(Ulrich et al., 1990),

(Dacorogna et al., 1993),(Andersen et al., 2001c), (Gencay et al., 2001a), (Gencay et al., 2001b)] except in section (4.4.2) where it is explained why 5-minute intervals are preferred to other frequencies. Section (4.4.2) also gives a remedy for how to overcome the problem of micro-structure noise while keeping most of the available data. Subsampling throws away most of the available data. For example, at the sampling frequency of 5 minutes per data point, there are only about 288 data points per day, while sampling at frequency of 1 minute per day there are about 1,440 data points, and sampling at frequency of 1 second per day there are about 86,400 data points.

The data used in this part of the chapter is five-minute GBP/USD closing prices for the period of 1 April 2009 - 1 April 2011 (in Greenwich Meantime) in their raw formats. The 1-minute observations are purchased from Olsen and Associates[1] and are further sampled at lower (5-minute and 1-day) frequencies.

Foreign exchange market is very unique in finance. First, it is a worldwide market with no business hour limitations, therefore each day has 24 hours worth of data and it is only inactive during weekends and national holidays. The first observation of the week arrives at 22:30 Greenwich Mean Time (GMT) on Sunday with the opening of the Asian markets and the last observation comes from the West Coast of the USA at about 22:30 (GMT) on Friday. Second, it is geographically diversified and spreads around the world and specifically this characteristic of the FX market has been the subject of several studies in which the effects of geographical locations on the autocorrelation functions of FX returns has been investigated [(Dacorogna et al., 1993), (Guillaume, 2000)]. Third, it has a significant high trading volume which shows it is one of the most liquid asset classes in the world and the bid-ask spreads are virtually zero and any new information is reflected in the price immediately. According to Bank of International Settlements, its volume on 5 September 2013 reached $5.3 trillion [2] which is equivalent to more than $757 for every single person on earth on that day[3]. Table (4.2) lists a summary of the data.

[1] Available at: http://www.olsendata.com. Person to contact with: Rakhal Dave, rakhal@olsen.ch

[2] In September 2013, the volume was on average $4 trillion. Available at: http://www.bis.org/publ/rpfx13fx.pdf

[3] Assuming that the population of earth is 7 billion.

Symbol	Observations	Start	End
1-minute GBP/USD	728,630	01/04/2009,00:01:00	01/04/2011,00:00:00
5-minute GBP/USD	145,750	01/04/2009,00:05:00	01/04/2011,00:00:00
Daily GBP/USD	625	01/04/2009	01/04/2011

Table 4.2: Each day has 24 hours. There are 60 1-minute intervals and 12 5-minute intervals in an hour which leads to 1440 1-minute intervals and 288 5-minute intervals within a day.

4.3 Intra-day Return

Similar to chapter 3, the simple rate of return (r_t) and the continuously compounded rate of return (R_t) on an asset are defined as:

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{4.2}$$

$$R_t = ln(P_t) - ln(P_{t-1}) = ln(r_t + 1) \approx r_t \tag{4.3}$$

Also recall from chapter 3 that estimating the mean and the variance of returns using a sample of daily observations:

$$\hat{\mu} = \frac{1}{T}\left(ln(P_T) - ln(P_0)\right) \tag{4.4}$$

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=0}^{T}\left((ln(P_t) - ln(P_{t-1})) - \hat{\mu}\right)^2 \tag{4.5}$$

If using daily observations, there are T observations that can be used to give us an estimate for σ^2. Imagine now instead of having price observations at the end of the day, prices are observed at the end of every minute and the market for an asset that is open all day long (for example an FX rate like GBP/USD). There are now $(24 \times 60 =) 1440 \times T$ observations for estimating the variance, σ^2. If prices were observed at 5-minute intervals, there would be $(\frac{1440}{5} =) 288 \times T$ observations and if prices were observed at the end of every hour, there would be $24 \times T$ observations available. Having more observations within a day (intra-day) at hand gives us more precise estimate than when using daily returns.

Since for estimating a monthly volatility 21 daily observations and for estimating an annual volatility 252 daily prices are used, similarly intra-day observations can also be used to give an estimate for daily volatility. Historical volatility is an unobserved variable but as the time horizon shortens volatility could "virtually" be treated as an observed variable and that is why when intra-day observations are used for estimating the daily variance, "realised" volatility replaces "historical" volatility. The idea behind using model-free measures of volatility based on realised returns goes back to (French et al., 1987) where daily observations are used for computing monthly realised variances. With the availability of intra-day data, it has become possible to improve the measure of the historical variance with the concept of realised variance, a concept which has been frequently used in the literature since (Andersen et al., 2001a). The use of intra-day data and notion of realised variance has given birth to the ongoing new research course of volatility modeling using intra-day data.

4.4 Realised Variance, RV

Assume that for an asset that is active 24 hours a day, there are m observations within a day. For 1-minute observations and a 24-hour trading window, $m = 1440$ and for 5-minute observations $m = 288$. The ith observation on day t can be denoted as $S_{t+\frac{i}{m}-1}$ and the closing price on day t is $S_{t+\frac{m}{m}-1} = S_t$. The ith 5-minute return is:

$$R_{t+\frac{i}{m}-1} = ln(S_{t+\frac{i}{m}-1}) - ln(S_{t+\frac{i-1}{m}-1}) \tag{4.6}$$

Having m return values, an estimate of the daily variance from the intra-day squared returns can simply be defined as:

$$\widehat{RV_{m,t}} = \sum_{i=1}^{m} R^2_{t+\frac{i}{m}-1} \tag{4.7}$$

There is a very important difference between equations (4.7) and (4.5) which is to do with not dividing the sum of squared returns by m in equation (4.7). If the sums in equation (4.7) is divided my m, then $\widehat{RV_{m,t}}$ would be an estimate for a 5-minute variance, whereas, the aim is to estimate the daily variance. Deleting the m gives a total variance for the 24-hour window.

	Mean	Stand. Dev.	Skew.	Kurt.	$Q(40)^a$	$Q^2(40)^b$
5-minute R	7.69e-07	0.0004486	-0.3065196	16.64147	304.0244	12422.2008
1-Day R	0.0001637	0.0061924	-0.0687485	3.759824	36.6128	94.8903

Table 4.3: Full summary statistics of GBP/USD returns from 01/04/2009, 00:05:000 through 1/04/2011, 00:00:00 at two different frequencies. **a:** Ljung-Box test statistics for up to fortieth order serial correlation in returns. **b :** Ljung-Box test statistics for up to fortieth order serial correlation in squared returns. $\chi^2_{40;0.1} = 51.805$ and $\chi^2_{40;0.01} = 63.691$ and $\chi^2_{40;0.005} = 66.756$

In equation (4.7), the mean of 5-minute returns is not subtracted since it is so small that it will not impact the variance estimate. The summary statistics of the data is presented in table (4.3).

The mean of the 5-minute GBP/USD is considerably smaller than its standard deviation. It is also interesting to note that the daily GBP/USD returns compared to the daily S&P500 returns show much less evidence of negative skewness (compare tables (3.5) and (4.3) which is typical of foreign exchange rates.

4.4.1 Properties of Realised Variance

Figure (4.3) illustrates two important stylised facts about realised variance which have been documented widely in the literature. The classic references on realised volatility include (Andersen et al., 2001b) and (Andersen et al., 2001a) with (Andersen et al., 2009) giving a thorough comparison between historical volatility and realised volatility.

Under suitable conditions (high liquidity and no microstructure noise), realised volatility is an unbiased and highly efficient estimator of return volatility as discussed in (Andersen et al., 2001b) followed by (Barndorff-Nielsen and Neil, 2002). The top-left panel of figure (4.3) shows the time series of daily realised GBP/USD variance calculated from intra-day squared returns using equation (4.7), whereas, the bottom-left part shows the daily close-to-close squared returns GBP/USD. Notice how much more irregular and noisy the bottom panel of the figure is in comparison with the realised variance which stems from the fact that realised variances are much more precise indicators (less noisy) for the daily variance than are daily squared returns. This is the first classified property of realised variance as an estimator for the volatility.

Figure (4.3) shows the GBP/USD realised variance series and the daily squared return series with their corresponding autocorrelation functions, in which the lag order

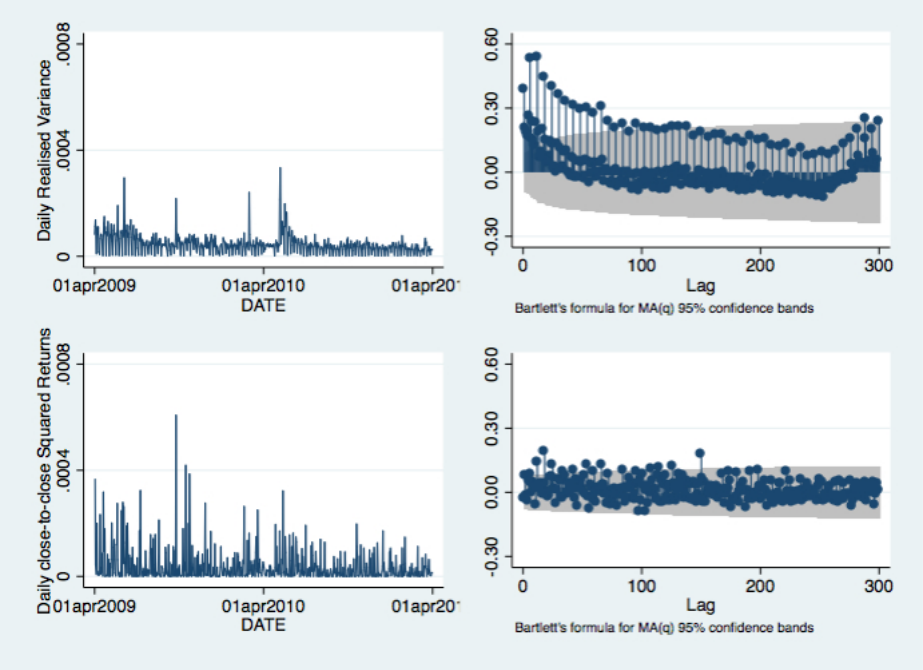

Figure 4.3: The top row of graphs show the daily realised variance and the autocorrelations of the daily realised variance. The bottom row of graphs show the daily close-to-close squared returns and their autocorrelation functions from the GBP/USD exchange rate

on the horizontal axis refers to the number of days. The difference in the degree of persistence in the top-right and the bottom-right parts of figure (4.3) is striking. This comparison resembles another important stylised fact of realised volatility: realised volatility is extremely persistent which indicates that volatility could be predictable for horizons longer than a month as long as intra-day data is used. This fact has been documented widely in the literature such as (Andersen et al., 2003), (Andersen et al., 2005) and (Corsi, 2009) where a simple framework for forecasting realised volatility is set up.

The top panel of figure (4.4) shows a histogram of the realised variance series and the bottom panel of figure (4.4) exhibits the histogram of the natural logarithm of realised variance. It can be seen that the level of realised variance series is strongly positively skewed with a right tail (leptokurtic), whereas, the logarithm of realised variance is much more closer to a normal distribution. Since realised variance is a sum of squared returns, it is understandable why the histogram of the realised variance is not close to a normal distribution and very useful that a natural logarithm function

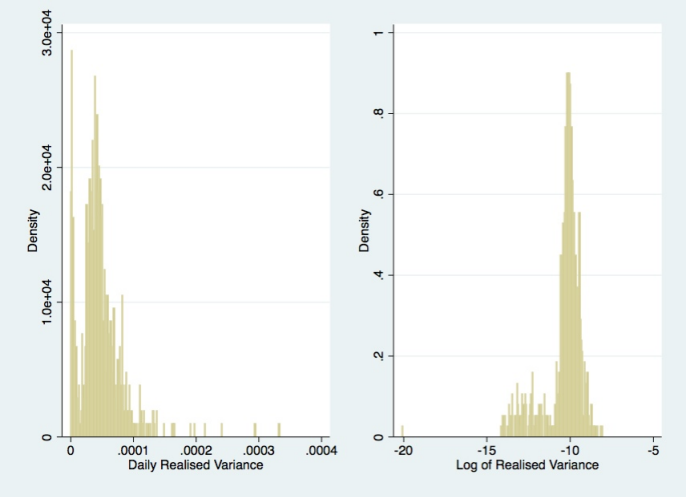

Figure 4.4: The histogram of the daily realised variance in the top panel and the histogram of the daily close-to-close squared returns in the bottom panel from the GBP/USD exchange rate.

	Mean	**Stand. Dev.**	**Skew.**	**Kurt.**	$Q(40)^c$
RV^a	0.0000469	0.0000352	2.521707	16.36266	1289.8305
Ln RV^b	-10.34129	1.082329	-1.545135	5.103155	2501.0507

Table 4.4: The sample covers April 1, 2009 through April 1, 2011. **a:** The distribution of realised variance RV_t **b:** The distribution of logarithmic realised variance, $\ln(RV_t)$. c: Ljung-Box test statistics for up to fortieth order serial correlation. $\chi^2_{40;0.1} = 51.805$ and $\chi^2_{40;0.01} = 63.691$ and $\chi^2_{40;0.005} = 66.756$

of the realised variance results in a distribution that is close to a normal distribution. The approximate log normal property of realised variance is another stylised fact:

$$ln(RV_{m,t}) \sim N\left(\mu_{RV}, \sigma^2_{RV}\right) \tag{4.8}$$

The statistics in the top panel of table (4.4) summarises the distribution of the realised variance, it is right-skewed and leptokurtic. In contrast, the skewness and kurtosis for the logarithmic realised variance, $ln(RV_t)$ shown in the bottom panel of the table, are much closer to the Gaussian distribution. The kurtosis has decreased by a factor of 3.2 and the skewness has decreased in absolute value by a factor of 1.63. The Ljung-Box statistics indicate strong serial correlation in the realised daily variances, which is in line with the significant Ljung-Box statistics for the squared (non-standardised) returns in the top panel of table (4.3). However, the $Q^2(40)$ statistic for

	Mean	**Stand. Dev.**	**Skew.**	**Kurt.**	$Q(40)^c$	$Q^2(40)^d$
R^a	0.0001637	0.0061924	-0.0687485	3.759824	36.6128	94.8903
Standardised R^b	0.0155267	0.8742985	0.0354926	2.703293	28.6123	44.4142

Table 4.5: **a:** The daily returns cover April 1, 2009 through April 1, 2011. **b:** The distribution of daily returns standardised by realised volatility. **c:** Ljung-Box test statistics for up to fortieth order serial correlation in returns. **d:** Ljung-Box test statistics for up to fortieth order serial correlation in squared returns. $\chi^2_{40;0.1} = 51.805$ and $\chi^2_{40;0.01} = 63.691$ and $\chi^2_{40;0.005} = 66.756$

the 1-Day return in table (4.3) (94.8903) is much smaller than the $Q(40)$ statistics in table (4.4). This reflects the fact that, relative to the daily realised variances, the daily squared returns are very noisy proxies for the variance (and hence for the volatility), since this noise brings down the persistence in the underlying volatility framework.

The fourth stylised fact is that the daily returns divided by the square root of realised variance are very close to an independently and identically distributed standard normal distribution:

$$\frac{R_t}{\sqrt{RV_{m,t}}} \sim^{i.i.d} N(0,1) \tag{4.9}$$

The statistics in table (4.5) refer to the two daily returns R_t (raw-return) from equation (4.3) in the top row and $\frac{R_t}{\sqrt{RV_{m,t}}}$ from equation (4.9) in the bottom row. The return series is approximately symmetric (slightly left-skewed) with a mean that compared to the standard deviation is negligible. The sample kurtosis shows more probability mass in the center and in the tails of the distribution relative to the normal (leptokurtic). The Ljung-Box test statistic indicates no serial correlation in returns, but strong serial correlation in squared returns. These results are consistent with section 3.4 and the extensive literature documenting fat tails and volatility clustering in returns, going back to at least (Mandelbrot, 1963).

The difference in the statistics of the standardised return is striking compared to results of the raw-return. First, the sample kurtosis and skewness indicate that the standardised returns are better approximated by a Gaussian distribution. Second, in contrast to the raw returns, the standardised return displays no evidence of volatility clustering. The Ljung-Box statistics for the standardised return squared is 44.4142 which is much smaller than the χ^2 critical value even at $\alpha = 10\%$.

Since $RV_{m,t}$ can only be computed at the end of day t, the result from equation (4.9) is not immediately functional for forecasting purposes. For instance if at 8:00 am on a Monday, a measure for the realised variance of that day is required, only the information from 22:30 GMT on Sunday with the opening of the Asian markets is available. There are two possible solutions to this problem. First, it is possible to forecast the quadratic variations of the return series at the preferred frequency and use these predicted values to calculate the realised variance. Recall from table (4.3) that the Ljung-Box statistics for the 5-minute return squared series is 12422.2008 which is 186 times greater than $\chi^2_{40;0.005} = 66.756$ and indicates strong persistence. The other solution would be to forecast the realised variance itself since from table (4.4), it is clear that $ln(RV)$ is also highly persistent. Section (4.5) is devoted to modeling the intraday return squared series and section (4.6) consider these two solutions for forecasting the realised variance.

4.4.2 Microstructure Noise

In an ideal situation, the daily variance for day $t+1$ can be calculated from:

$$RV_{m,t+1} = \sum_{i=1}^{m} R^2_{t+\frac{i}{m}} = \sum_{i=1}^{m} \big(ln(P_{t+im}) - ln(P_{t+(i-1)m})\big)^2 \tag{4.10}$$

However when dealing with intraday and high frequency data observed price series $P_t^{Observed}$ is different from the actual underlying price series $P_t^{Unobserved}$ due to the process of price formation and market microstructure noise such as bid-ask bounces and discreteness of price changes. If P_t^O is the observed price series and P_t^U is the unobserved price series, then:

$$P_t^O = P_t^U + e_t \tag{4.11}$$

and if R_t^O is the observed return series and R_t^U is the unobserved return series then:

$$R_t^O = R_t^U + e'_t \tag{4.12}$$

where e_t and e'_t are assumed to be the microstructure noise series in the price and the return series. Inserting equation (4.12) back into equation (4.10), for the observed return series:

$$\mathrm{RV}_{m,t+1} = \sum_{i=1}^{m} (R^{O}_{t+\frac{i}{m}})^2$$

$$RV_{m,t+1} = \sum_{i=1}^{m} \left(R^{U}_{t+\frac{i}{m}} + e'_{t+\frac{i}{m}} \right)^2 \tag{4.13}$$

And for the unobserved return series:

$$RV_{m,t+1} = \sum_{i=1}^{m} (R^{U}_{t+\frac{i}{m}})^2 \tag{4.14}$$

Clearly the realised variance computed from equation (4.13) is higher than the realised variance estimated from equation (4.14) and hence the realised variance computed by using sum of return squared series is noisy and different from actual realised variance.

The presence of microstructure noise in high-frequency financial data makes the realised variance estimator unreliable. Among others, (Ulrich et al., 1990), (Dacorogna et al., 1993) provide good surveys on the issues one faces when dealing with intraday observations. One of the first articles in the literature that dealt with estimating the volatility in the presence of microstructure noise goes back to (Zhou, 1996). Others include (Andersen et al., 2003) and (Hansen and Lunde, 2006) offer a more general and thorough frameworks for understanding the dynamics of microstructure noise. This area of research is very much active and there are different solutions for computing realised variance while dealing with the microstructure noise. The following sections will provide two widely used solutions.

4.4.2.1 Subsampling, RV_{SUB}

Since the presence of microsutructure noise increases as the sampling frequency for intraday returns increases, perhaps the simplest way to overcome the microstructure noise problem is to construct a grid of intraday returns that are sampled less frequently. For instance instead of using 1-minute intervals, larger intervals could be used, maybe 5-minute intervals or 20-minute intervals which are the two most frequently used in the literature. There is a trade off between using larger intervals and useful data, so one shouldn't use larger intervals without optimising (minimising) the microstructure noise. Here we show how this is done.

Assume that the unobserved price series and hence the unobserved return series have become observable, the higher the sampling frequency (smaller intervals), the more precise the estimation of the realised variance (and hence volatility) will be. However, in practice the unobserved return series are not observable, and as sampling frequency increases, the noise also increases and it decreases the chance of achieving an unbiased estimator. This is the typical variance-bias trade-off problem cited in the literature [(Hansen and Lunde, 2006) and (Bandi and Russell, 2008)]. So to overcome this problem, lower frequencies could be used at the expense of throwing away information. The choice of sampling frequency depends on the asset and even the optimal sampling frequency for an asset changes over time. For very liquid assets with high trending volumes usually sampling frequencies are between 1-minute and 5-minute intervals, whereas, for less liquid assets sampling intervals are much larger.

To choose the correct sampling interval, a graphical tool called the variance signature plot has become very popular since (Andersen et al., 2000). The process consists of three steps. First computing $RV_{m,t}$ for different sampling frequencies (m) from the finest grid available to 60 minutes (m often going from 1 to 60 minutes, for liquid assets). Second, the average of RV across days are plotted on the vertical axis against all the sampling frequencies on the horizontal axis. Finally, the smallest sampling frequency at which the average RV stabilises and does not change much for larger values than this number, is chosen as the optimal sampling frequency.

Figure (4.5) presents the variance signature plots for GBP/USD for two years, the top plot uses information from 01/04/2009 to 01/04/2010 (hence year1) for averaging realised variances and the bottom plot uses information from 01/04/2010 to 01/04/2011 (hence year2). For $m = 1$, the average realised variance is calculated using 1-minute returns and it is averaged across 01/04/2009-01/04/2010 for the top row graph and averaged across 01/04/2010-01/04/2011 for the bottom row graph in figure (4.5). For $m = 2$, 2-minute returns are used for constructing the average realised variance, and so forth. There is little difference between the top row and the bottom row graphs as to which sampling frequency (m) is the optimal. 5-minute sampling frequency is chosen the optimal sampling frequency at which the balance between variance and bias is met and this is the reason in figure (4.3), 5-minute return have been used.

From figure (4.5), realised variances based on moderate frequencies appear to be somewhat unbiased, however at higher frequencies realised variances are not reliable and

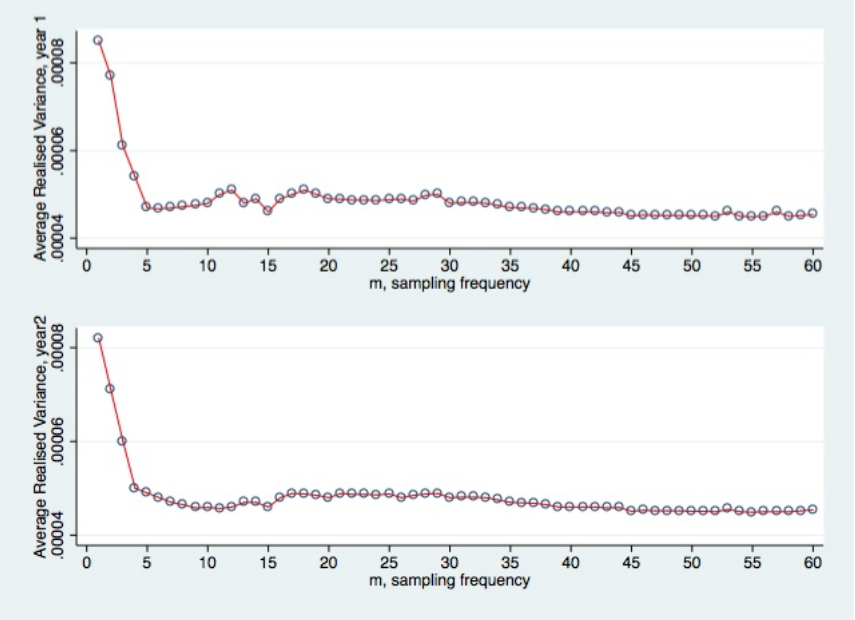

Figure 4.5: Volatility signature plots for GBP/USD from 01/04/2009 to 01/04/2011 divided over two years. Average realised variance on the vertical axis and sampling frequencies on the horizontal axis.

transfer the microstructure noise. For example at the 1-minute frequencies, the realised variance is about 1.7 times larger that the realised variance at 5-minute frequencies.

4.4.2.2 The Average Realised Variance Estimator, RV_{AVE}

When using RV_{SUB} considerable amount of information is thrown away, which is wasteful. Another simple way which has proved to be performing well among other realised variance estimators is proposed by (Zhang et al., 2005) and is called the "average realised variance" estimator.

The process is as follows. Having used the variance signature plot to find the optimal sampling frequency leads to $m = 5$, hence 5-minute returns are chosen to compute the realised variance. Since the original data is in 1-minute intervals, it is possible to calculate 5 different (overlapping) RV_{SUB} estimators. The first $RV_{1,SUB}$ estimator starts the 5-minute interval at midnight. The second $RV_{2,SUB}$ estimators starts from one minute past midnight, the third $RV_{3,SUB}$ estimators starts from the two minutes past the midnight and so on until $RV_{5,SUB}$ estimators starts from 4 minutes past midnight. Therefore, 1-minute returns have been used to calculate 5 RV_{SUB} estimators

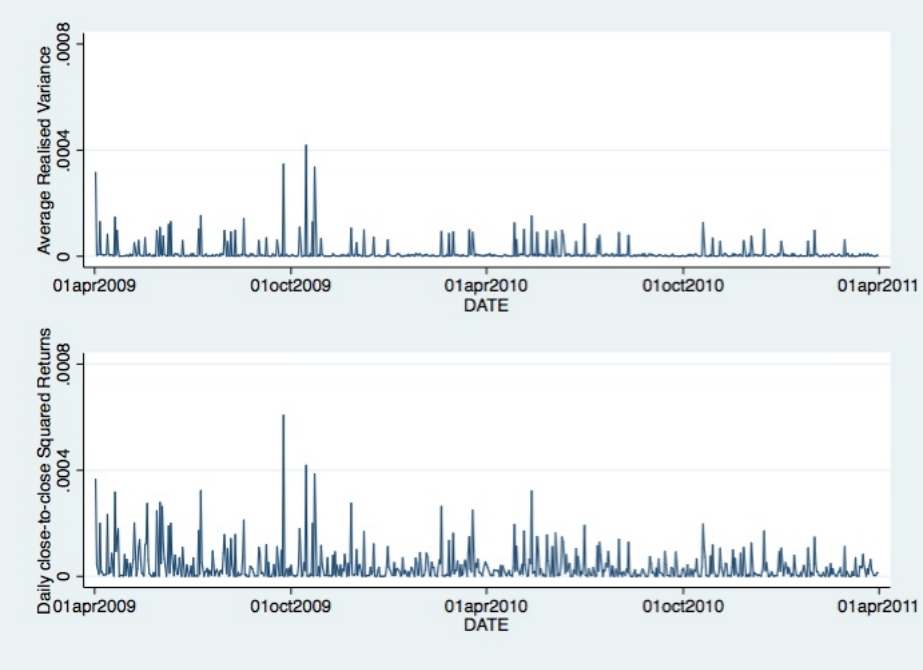

Figure 4.6

	Mean	**Stand. Dev.**	**Skew.**	**Kurt.**	$Q(40)$	$Q^2(40)^d$
$ln(RV_{AVE})^a$	-12.96577	1.093	0.251	3.29	5983.2	
Standardised R^b	0.0027	0.993	0.032	2.9032	19.4213	23.3204

Table 4.6: a: The distribution of logarithmic average realised variance, $\ln(RV_t)$. **b:** The distribution of daily returns standardised by average realised volatility, $\frac{R_t}{\sqrt{RV_{AVE,t}}}$. **c:** Ljung-Box test statistics for up to fortieth order for serial correlation. **d:** Ljung-Box test statistics for up to fortieth order serial correlation in squared returns. $\chi^2_{40;0.1} = 51.805$ and $\chi^2_{40;0.01} = 63.691$ and $\chi^2_{40;0.005} = 66.756$

at the 5-minute frequency. By averaging the 5 $RV_{i,SUB}$ estimators the so-called Average RV estimator is defined as:

$$RV_{AVE,t} = \frac{1}{m}\sum_{i=1}^{m} RV_{i,SUB,t} \tag{4.15}$$

Figure (4.6) represents the $RV_{AVE,t}$ in the top row and the daily close-to-close squared returns in the bottom row for GBP/USD. The top panel of figure (4.6) compared with the top panel of figure (4.3), shows that $RV_{AVE,t}$ is a much less noisier estimator that $RV_{SUB,t}$.

It is obvious from comparing table (4.6) with tables (4.5) and (4.4) that $ln(RV_{AVE})$ is much more persistent than $ln(RV_{SUB})$ based on 5-minute returns. Also the standardised returns appear remarkably Gaussian. These results are all in line with (Andersen et al., 2000), (Zhang et al., 2005), (Hansen and Lunde, 2006) and (Bandi and Russell, 2008).

4.5 Modeling Intra-Day Quadratic Returns

As discussed before, $RV_{m,t}$ (either $RV_{AVE,t}$ or $RV_{SUB,t}$) can only be computed at the end of day t, i.e., at the early minutes after the midnight, a measure for the realised variance for that day is not available. However, there are two possible ways to forecast the persistent realised variance. The first one is to forecast the realised variance directly, which seems very likely given how persistent $RV_{AVE,t}$ and $RV_{SUB,t}$ are. The second way would be to forecast of the quadratic return series at the finest grid and then use the forecasted intraday return squared series to calculate the realised variance. Recall from table (4.3) that the Ljung-Box statistics for the 5-minute return squared series is 12422.2008 which is 186 times greater than $\chi^2_{40;0.005} = 66.756$ and indicates the strong persistence in squared returns. The two scenarios are both considered here, but first a framework for forecasting intraday return squared series needs to be set up.

Figure (4.7) plots the price series of GBP/USD for over two years, while figure (4.8) represents the 5-minute returns, 5-minute squared returns and their autocorrelation functions. The autocorrelation function for GBP/USD 5-minute returns for up to 24 hours is plotted in top-right part of figure (4.8). Negative autocorrelation is observed for up to 10 minutes and for few longer lags which could be because, traders might have different opinions about the impact of news on the direction of price (and hence the sign of the return). The summary statistics of 5-minute GBP/USD returns are presented in table (4.3).

Looking at the bottom-right part of figure (4.8), one can see periods of "$\bigcup$" shapes which appear at every 288 lags. There are 288 5-minutes in one day and what this graph shows, is the important and established evidence of intraday seasonality. This phenomenon has been reported widely in the literature going back to (Ulrich et al.,

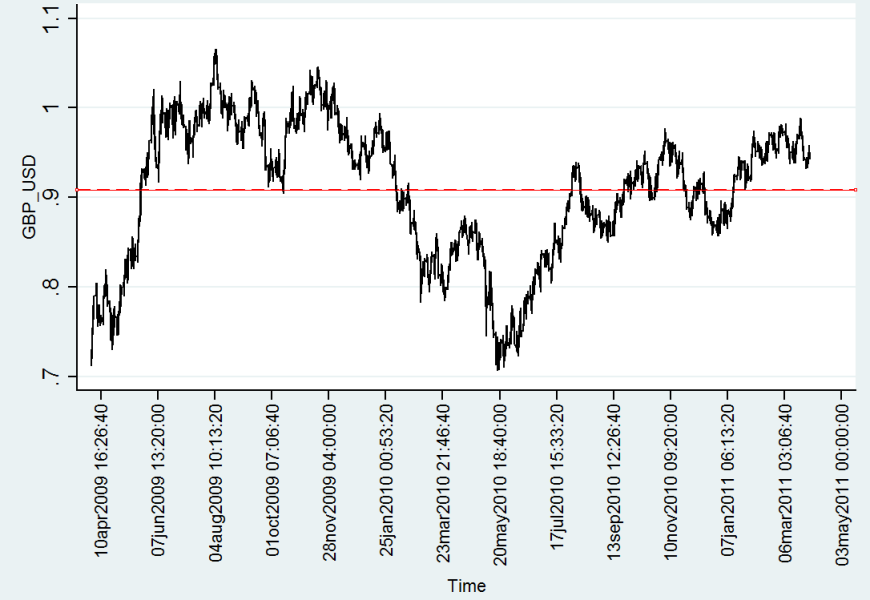

Figure 4.7: The price series of GBP/USD spot exchange is plotted against time, from April 2009 till March 2011.

1990), (Dacorogna et al., 1993), (Olsen et al., 1997), (Andersen and Bollerslev, 1997b) and (Goodhart and O Hara, 1997).

As mentioned, FX markets do not close and are open 24 hours and hence the intraday patterns reflect the ebb and flow of trading activity across the world. The squared returns which are the building blocks of volatility fluctuate dramatically over day. The average squared returns over the 5-minute intervals are presented in figure (4.9) which reveals a pronounced difference in the values of squared returns over the day[1]. The squared returns values reach lows such as 2.4×10^{-5} at 22:55:00 GMT (interval 276), and highs of around 8.2×10^{-4} at 14:30 GMT (interval 174). This pattern is closely linked to the cycle of market activity in the various financial centers around the globe. Table (4.7) summarises the important trading intervals.

[1]The average squared returns over 5-minute intervals are calculated as follows: The squared return at 00:05:00 for each day over two years are averaged, then the squared return at 00:10:00 for each day for two years are averaged, and so forth.

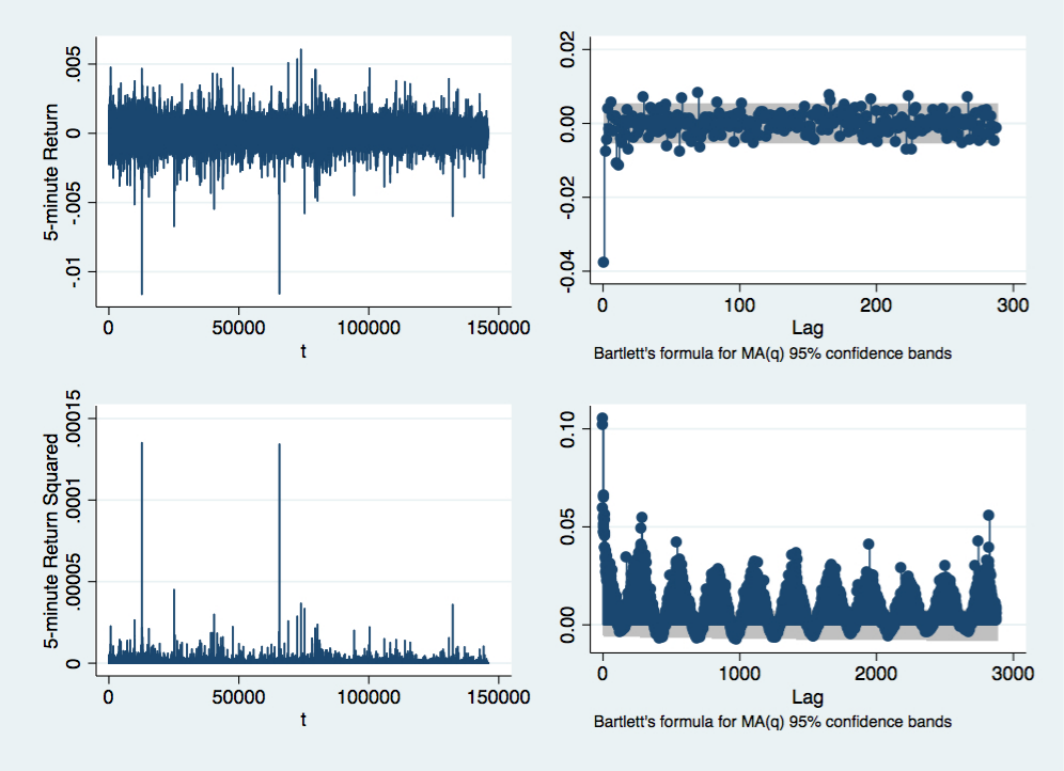

Figure 4.8: The top part: 5-minute GBP/USD returns and the autocorrelation coefficients up to lags 288 (24 hours). Negative autocorrelation is observed up to a time lag of 10 minutes for the return series. The bottom part: 5-minute GBP/USD squared returns up to lag 2880 (10 days).

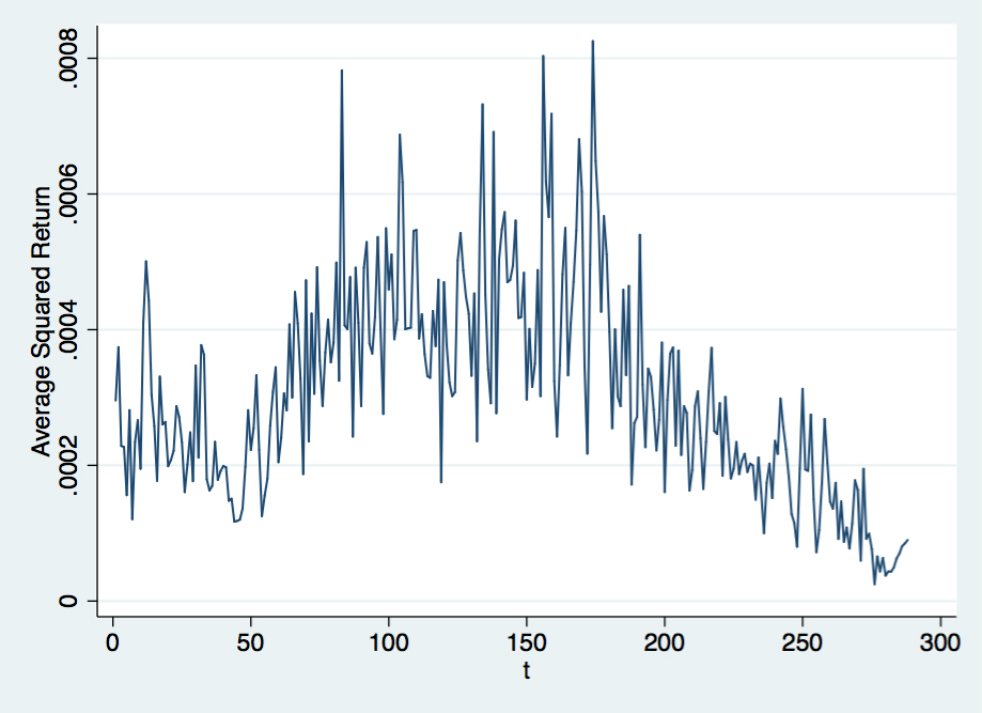

Figure 4.9: Intraday squared returns averaged over 5-minute intervals for GBP/USD.

The activity starts at a relatively high level followed by a slow decay between intervals 45 and 60 which corresponds to the lunch hour in the Asian markets. By opening European markets around 07:00 GMT squared returns increase followed by a decrease at around 11:30 GMT. However, it increases significantly by opening of the US markets at around 13:00 GMT. The squared returns reach their maximum at 14:30 while the European and the US markets overlap. By closing hour of the European market, around 17:00 GMT, the activity declines until it goes up slightly at 22:00 GMT with opening hours of Sydney. The robustness of these intraday ups and downs are confirmed in similar studies such as Dacorogna et al. (1993), Andersen and Bollerslev (1997b), (Andersen and Bollerslev, 1997a) and Andersen and Bollerslev (1998b).

The behavior of a time series is called seasonal if it shows a periodic structure in addition to less regular movements. The pronounced systematic ups and downs in figure 4.8 is an indication that usual modeling of intraday quadratic variation would dangerous since they simply cannot accommodate strong regular cyclical patterns. One of the most frequent situations that arises from analysing high frequency financial data is the fact that, long memory and seasonality are both evident in the dynamics of intra-

Trading Event	Time (GMT)	Interval
Tokyo, open	midnight	0
Hong Kong, open	1:00 am	12
Asia lunch hour	3:55 am-5:10 am	45-60
Frankfurt, open	7:00 am	84
London, open	8:00 am	96
European lunch hour	11:30	138
New York, open	1:00 pm	156
Chicago, open	2:00 pm	168
Maximum reached	2:30 pm	174
London, close	5:00 pm	204
Sydney, open	10:00 pm	268
Minimum reached	10:55	276

Table 4.7: Trading sessions and events according to GMT time.

day data. The presence of seasonality and long memory at intraday levels could be problematic since, the presence of seasonalities in a long memory process may obscure the underlying low-frequency dynamics. The periodic component pulls the calculated autocorrelations down, giving the impression that there is no persistence other than particular periodicities and hides the long memory feature of the data. Also, strong intraday seasonalities (like the ones observed in figure 4.8) may induce distortions in the estimation of volatility models which are based on the dynamics of squared returns, such as realised volatility.

In order to extract intraday seasonality a model of intraday returns which are widely used in the literature since (Andersen and Bollerslev, 1997a) and (Gencay et al., 2001a), is used:

$$r_t = v_t s_t \epsilon_t \tag{4.16}$$

where r_t is the raw return, v_t is the long term volatility, s_t is the seasonal volatility and ϵ_t is the identically and independently distributed innovations [1]. Squaring both sides of equation 2, taking in natural logarithm and dividing both sides by two leads to:

[1] On the left hand side of equation 4.16 r_t can have positive or negative values. On the right hand side of the equation v_t and s_t can only have positive values and ϵ_t can have positive or negative values.

$$log|r_t| = log|v_t| + log|s_t| + log|\epsilon_t| \tag{4.17}$$

An obvious route would be to set up a framework that deals with filtering out the underlying intraday seasonalities first. Modeling intraday seasonalities began by the works of (Ulrich et al., 1990) and (Dacorogna et al., 1993) in which the sum of three polynomials corresponding to the distinct geographical locations of the market are used. However, this method didn't become very popular as depending on the asset the effects of the geographical locations change. There are other averaging and smoothing methods ((Taylor, 2010)) in the literature but the most dominant one has been the Fast Fourier Transform (FFT) widely used in the literature starting from (Andersen and Bollerslev, 1997a), (Andersen and Bollerslev, 1998b) and (Phillips, 1999). Among others (Martens et al., 2002) give a comparison between available methods for filtering. They specifically compare how much filtering the squared returns help in improving the forecast performance. They use different combinations of FFT and GARCH models and conclude that filtering the intraday data with FFT first and then using an estimator to forecast the volatility seems to be the most reasonable framework.

The estimation of the FFT (similar to the one used in(Dacorogna et al., 1993)) is dependent on several decisions (selecting the interaction terms conditional on the shape of the periodic trends) which are based on the researcher's view of the data. Also, in order to remove the short-term intraday periodicities more parameters are required which is then not desirable.

A simple wavelet method similar to the one used in (Gencay et al., 2002) can be utilised for extracting intraday seasonality that does not depend on a particular model selection criterion or parameter choices and can be used on any high-frequency data. This method is based on a wavelet multi-scaling approach which decomposes the data into low and high frequency components of data by applying a discrete wavelet transform. This method was explained in section 2.4.5.

Equation 4.17 provides an additive decomposition of $log|r_t|$. In order to break $log|r_t|$ into $log|v_t|$ and $log|s_t|$ we have applied wavelets (similar to the ones used in (Gencay et al., 2001a) (Nason, 2008)) on $log|r_t|$. The type of the wavelet transform used is a Maximal Overlap Discrete Wavelet Transform (MODWT) which is used mostly for extracting seasonalities from high frequency data. A nine level MODWT is utilized to

	Return	Squared Return	Absolute Return
5-minute filtered GBP/USD	1.48	9.46	9.33
5-minute GBP/USD	1.3	7.94	9.13
1-Day GBP/USD	1.39	2.15	1.9

Table 4.8: The summary of the Modified R/S test statistic for the 5-minute return, 5-minute filtered return, daily return and their squared and absolute terms. The Critical Values are as follows: $90\% : [0.861, 1.747]$, $95\% : [0.809, 1.862]$, $99\% : [0.721, 2.098]$

decompose the $log|r_t|$ at the $5-$minute frequency. The Daubechies least asymmetric family of wavelets ($LA(8)$) was utilized in maximal overlap discrete wavelet transformation.

The filtered returns are defined as:

$$log(f_t) = logf_{t,1} + logf_{t,2} + logf_{t,3} + ... + logf_{t,9}. \tag{4.18}$$

where $logf_t$ corresponds to intra-day seasonal volatility (s_t) and high frequency components of the innovations (ϵ_t) obtained from MODWT details. As a result of this decomposition the series left corresponds to $log|v_t|$.

To explain the right hand side of Equation (4.18) and using the jargons used in chapter 2: $logf_{t,1}$ is the first detail in MODWT and it associated with $5 \times 2^0 = 5-$ minute changes which shows the highest frequency part of the innovations. Similarly, $logf_{t,2}$ is the detail 2 and deals with $5 \times 2^1 = 10-$ minute periodicities and is the second highest frequency part of the innovations. The highest detail $logf_{t,9}$ contains 1280 minutes (approximately 21 hours) periodicities and since there are 1440 minutes in a day, it corresponds to intraday dynamics. The filtered absolute returns, therefore, are free from any intra-day periodicities and innovations.

In the bottom part of figure 4.10 the autocorrelations of the filtered $5-$minute absolute returns are presented. Now that the intraday seasonalities have been removed the slow decaying rate of the autocorrelations have become more clear. It is clear from table 4.8 that the 5-minute squared returns and the 5-minuted filtered squared returns show signs of long range dependence. The Ljung-Box for the filtered return squared is 9.955×10^6 which is much larger than the Ljung-Box for the 5-minutes return squared which is 94.8903.

Using the three estimators described in chapter 3, the GPH, the Whittle and the

WMLE estimators parameter d is estimated for the 5-minute filtered return squared in ARFIMA(0,d,1) and the results are presented in table (4.9):

	GPH	Whittle	WMLE
5-minute Filtered Return Squared	0.3390	0.3759	0.3816
	(0.0019)	(0.0051)	(0.0025)

Table 4.9: $\hat{d}$ and their standard errors in brackets for the 5-minute absolute filtered returns using the GPH, the Whittle and the WMLE estimators in ARFIMA(0,d,1).

The autocorrelations of the filtered 5-minute absolute returns for GBP/USD is plotted in figure 4.10 which exhibit the autocorrelation coefficients of the unfiltered and filtered squared returns. The comparison between the periodic and hyperbolic decay rates are clear between the two series.

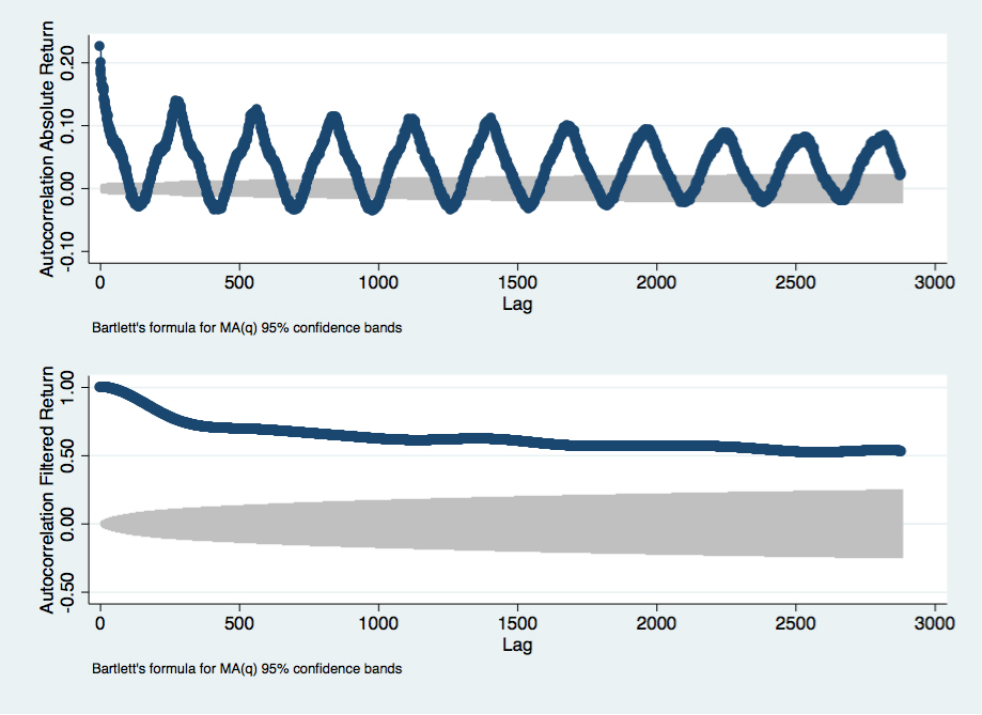

Figure 4.10: Sample autocorrelation of the 5-minute GBP/USD return and the MODWT filtered 5-minute absolute returns of GBP/USD spot exchange rate from 1, April 2009 to March 2011

4.5.1 Temporal Aggregation at Intraday level

Similar to section (3.6) a natural question arises whether some sampling intervals are superior relative to others in terms of estimating the persistence parameter and whether temporal aggregation changes the degree of long-range dependence. The comparison between equations (3.73) and (3.37) and also the results from the simulations show that the decay rates of the aggregated ARFIMA(0,d,0) and ARFIMA(0,d,0) are the same (τ^{2d-1}) but whether seasonality can cause a problem or whether intra-day data inherently do not fall into this category is an important question. As widely noted by (Zhou, 1996), (Olsen et al., 1997), (Andersen and Bollerslev, 1998b) and (Dacorogna et al., 2001) the results regarding the degree of persistence in the return squared or the volatility of intra-day data can be puzzling and often contradicts the works of (Drost and Nijman, 1993) in terms of the aggregation results of the volatility models. There have been some contradictory evidence in the literature regarding the estimates of the long-memory parameter when working with intraday data. For example (Mcmillan and Speight, 2008) document results through estimating d for different currencies at 30-minute, 1-hour, 2-hour and 4-hour frequencies and find that temporal aggregation apart from the upward bias discussed in section (3.6) does not contaminate the estimations. While (Bollerslev and Wright, 2000) find different volatility measures to be performing differently when they are temporally aggregated, for example return squared and absolute return as volatility measures have shown signs of being noisy estimators hence leading to a downward bias when temporally aggregated.

There could be three possible explanations that help solving the puzzle. Perhaps the most basic explanation that often gets ignored is that the theoretical explanation in section (3.6.2) is not valid when strong intraday periodocities are present. After all x_t in equation (3.67) is assumed to follow an ARFIMA model, and not an ARFIMA model contaminated with strong periodocities which is the case for almost all intraday financial time series, either exchange rate or indexes. That is why, the filtered GBP/USD 5-minuted squared returns have been sampled at 30-minute, 60-minute, 10-hour, 1-day and 5-day intervals and the long-memory parameter d has been estimated using the two GPH and WMLE estimators in ARFIMA(0,d,1). The results are presented in table (4.10):

	GPH	WMLE	n
5-minute	0.3390 (0.0019)	0.3759 (0.0051)	145750
30-minute	0.3516 (0.0089)	0.3815 (0.0067)	24291
60-minute	0.3418 (0.0127)	0.3560 (0.0076)	12145
10-hour	0.3476 (0.0467)	0.3634 (0.0260)	1214
1-Day	0.3269 (0.0508)	0.3345 (0.0307)	627
5-Day	0.2582 (0.1375)	0.2612 (0.0982)	125

Table 4.10: d for the filtered 5-minute, 30-minute, 60-minute, 10-hours, 1-day and 5-day sampling intervals using the GPH and the WMLE estimators in ARFIMA(0,d,1).

Starting with 5-minute intervals, there is clear sign of long-rage dependence both with GPH and WMLE estimators. The $\hat{d}$ increases at 30-minute intervals which was expected due to the upward bias from aggregation. However at 1-hour sampling interval there is a decrease followed by a slight increase at the 10-hour sampling interval. It maybe that somewhere between 30-minute and 60-minute sampling intervals $\hat{d}$ decreases. From table (4.10), it is clear that at 30-minute sampling interval temporal aggregation has led to an upward bias in the estimation of d, however at 10-hour and 1-day sampling interval $\hat{d}$ has decreased. This has two explanations either the MRA (Multi-Resolution Analysis) filtering hasn't been powerful enough to filter out all seasonalities and there is still some periodocities left in the data that distort the autocorrelations, or recall from equation (3.78), that as $h \to \infty$ $\hat{d}$ doesn't depend on h (sampling interval). Since $10-hours$ has 120 $5-minute$ and 1-day has 288 $5-minute$ intervals the latter explanation seems logical. Furthermore, the bottom part of figure (4.10) doesn't show any signs of seasonalities so it is more likely that seasonalities have been removed from the data. At 5-day sampling interval, there is only 125 data points left which explains why it has such a high standard error. These results are consistent with the results from (Mcmillan and Speight, 2008).

The other two possible explanations out of the three that might help solving the puzzle of temporal aggregation at intraday data are as follow. First, the discussion by (Bollerslev and Wright, 2000) that not not all volatility estimators follow the same

rule when it comes to temporal aggregation. Second, perhaps the reason why there are contradictory results in the literature is because not all the studies use exactly similar sampling intervals and do not start from exactly the same sampling interval. This rather basic decision as to what the smallest sampling frequency is, can be crucial. As discussed above from equation (3.78), as $h \to \infty$ $\hat{d}$ s do not depend on the sampling frequency. So how long this effect will take to happen, depends on what the first sampling frequency is, what sampling frequencies are used in temporal aggregation and how low the last sampling frequency is. It is obvious that even if for a single time series, different combinations of aggregation could result in different outcomes in terms of the direction of $\hat{d}$ s.

4.6 Modeling Realised Variance: Scenarios A & B

Since $ln(RV_{AVE})$ proved to be more persistent than $ln(RV_SUB)$, an ARFIMA(0,d,2) is chosen for modeling $ln(RV_{AVE})$. The one-day-ahead return variation forecast, is of huge interest for practitioners. As discussed earlier, a practitioner in the morning might need a volatility forecast for that day. Two scenarios have been considered here. First is using $ln(RV_{AVE})$ for yesterday and forecast a volatility estimation for today and the second is to use the intraday returns of past days to forecast the intraday returns for today and then calculate the realised variance based on these forecasted returns. We call this RV_R.

The standard approach to study the forecast performance of volatility models is by regressing the true integrated variance on day $t+1$, on a constant and various estimator forecasts, the so-called *Mincer-Zarnowitz* regression (Mincer and Zarnowitz, 1969). For example, for $ln(RV_{AVE})$, forecasting performance is evaluated by running the following regression:

$$IV_{t+1} = \alpha + \beta \times RV_{AVE,t+1|t} + \epsilon \tag{4.19}$$

Where $RV_{AVE,t+1|t}$ is the one-day ahead forecast of integrated variance from day t to day $t+1$ using the AR(1) prediction. Equation (4.19) regresses the true realised variance IV_{t+1} from day $t = 1$ on a constant and the variance forecast using RV_{AVE} estimator. If the RV_{AVE} estimator performs well, the forecast should be unbiased and

the forecast error small. In other words, $\alpha = 0$ and $\beta = 1$ and the R^2 of the regression is close to 1. Thus, the null hypothesis of $H_0 : \alpha = 0$ and $H_0 : \beta = 1$ against the alternatives $H_1 : \alpha \neq 0$ and $H_1 : \beta \neq 1$. The results are presented in the table below:

	Constant	RV_{AVE}	RV_R	Rˆ2
RV_{AVE}	-0.55 (0.003)	1.268 (0.089)		0.941
RV_R	-0.063 (0.003)		1.567 (0.013)	0.895

Table 4.11: Out-of sample Mincer-Zarnowitz regression results for RV_{AVE} and RV_R

It can be seen from the table above that RV_{AVE} is a better performance, although RV_R is only marginally worse.

5

Conclusion

This thesis is concerned about estimating and measuring the volatility and quadratic variation at diffeent scales. The long-memory feature of financial returns is of a relavant theme that is dealt with under temporal aggregation at both intra-day and daily levels.

This thesis is divided into three sections.

Chapter two of the thesis, documents the wavelet multiresolution analysis as an alternative method to the Short Time Fourier Transform that does not contradict with the Heisenberg's Uncertainty Principle. This chapter defines wavelet function, filters and coefficients. It explores the usefulness of multiresolution analysis for decomposing the time series into several sub-series which is used in chapter 4, and the concept of wavelet variance is defined which is used in chapter 3.

Chapter three with the use of daily S&P500 return for a decade, lists important empirical properties of the daily returns. Long memory processes are defined, and Autoregressive Fractional Integrated Moving Average (ARFIMA) models as a good candidate for modelling the data are proposed. Several testing and estimating methods are explored, such as the modified rescaled range (R/S) test, the Geweke-Porter-Hudak log periodogram regression (GPH). The maximum likelihood estimation for estimating d and the computational aspect of it gives its place to the approximate versions such as the Wavelet Maximum Likelihood Estimator. With the use of simulation the power of the three tests are compared. These are the GPH test, the Whittle estimator (another method for the approximate maximum likelihood estimation), and the WMLE. GPH tests seems to be the one that gives closer estimates to the true values of d but with higher standard errors.

The dillema of calculating risks at different horizons is explored and as a consequence temporal aggregation and its effects on estimating the parameters in ARFIMA processes are studied. We show theoretically and by simulations that temporal aggregation does not change the decay rate of a long-range dependent process for when $0 < d < 1/2$ but it will result in an upward bias in estimating the parameter d in ARFIMA(0,d,0). One way to avoid this is to increase the short memory parameters of an aggregated ARFIMA(0,d,0). This chapter ends with calculating $\hat{d}$ for squared daily returns, squared 5-day returns, and squared 10-day returns in ARFIMA(0,d,0) and ARFIMA(0,d,1) using the GPH method.

Chapter four of the thesis, analyses the five-minute Foreign Exchange (`GBP/USD`) returns for two years. With the use of intraday data, after listing the empirical properties of the returns, the concept of realised variation (RV) is defined and it is contrasted with the unobserved estimation of the volatility. With the existence of microstructure noise, two ways of calculating the realised variance is proposed, these are the RV_{SUB} (subsampling realised variance) and the RV_{AVE} (the averaging realised variance). RV_{AVE} proves to be more capable to minimise the microstructure noise effect when calculating the realised variance. Both measures of realised variation and the 5-minute squared returns show signs of long-range dependence and ARFIMA models prove to be good model choices.

Similar to chapter three, the effects of temporal aggregation at intra-day levels are also monitored. There have been some contradictory evidence in the literature regarding the estimates of the long-memory parameter when working with intraday data. The direction of the decrease or the increase of long-range dependence (d) seems to be different among intraday studies. Claims about whether $\hat{d}$ increases or decreases when the sampling frequency increases have been made and most of them are contradictory. We propose three possible explanations. First, since at intraday levels strong intraday periodocities are present, estimating long-memory parameters can result in misleading outcomes. Only after removing the seasonality, $\hat{d}$ s can be reliable. Second, because of the existence of microstructire noise at intraday levels, different volatility measures perform differently when they are temporally aggregated (first proved in Bollerslev200081) and since absolute returns or squared returns have been used as estimators for volatility, the results will be different. The third explanation, which according to our knowledge has not been mentioned anywhere else, goes back to when $h \to \infty$ $\hat{d}$ s do not depend

on the sampling frequency as discussed in (3.78) and since not all studies use exactly similar sampling intervals and do not start from exactly the same sampling interval, getting different results are not avoidable.

The main contribution of this thesis lies in its approach to the effects of temporal aggregation when the original data follows an ARFIMA process. This is an active area of research and as for future research, we intend to focus our attention on mature and emerging indexes and study the effects of temporal aggregation at these two different environments.

Appendix A

Proof for equation 4.3:

We know that:

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{1}$$

And we need to prove that continously compounded return:

$$R_t = ln(P_t) - ln(P_{t-1}) \tag{2}$$

Since it is the continously compounded return (R_t) that we are concerned, these are our assumptions:

1. There is an initial amount; A invested.
2. The rate of return is R; could be annual interest rate.
3. Units of time t; could be years.
4. Amount remaining after t units of time, B.
5. Number of times interest is paid within t units of time, n.

For example if the interest is paid twice a year, $n = 2$ and $t = 1$ year.

What we are interested in is to simplify $r = \frac{B-A}{A}$.

First we need to find what B is. For simplicity we first assume that the interest is paid once a year, so $n = 1$ (interest is paid annualy).

At $t = 0$, B=A

At $\underline{t = 1}$, AR is the profit added to A, hence:

$$B = A + AR = A(1 + R) \tag{3}$$

At $\underline{t = 2}$, $R(A + AR)$ is the profit added to $A + AR$, hence:

$$B = A + AR + AR + AR^2 = A + 2AR + AR^2 = A(1 + 2R + R^2) = A(1 + R)^2 \tag{4}$$

At $\underline{t = 3}$, $R(A + 2AR + AR^2)$ is the profit added to $A(1 + R)^2$, hence:

$$B = A+2AR+AR^2+AR+2AR^2+AR^3 = A+3AR+3AR^2+AR^3 = A(1+3R+3R^2+R^3) = A(1+R)^3 \tag{5}$$

For $\underline{t = T}$:

$$B = A(1 + R)^T \tag{6}$$

If $\underline{n = 2}$ (interest paid twice a year) then at $t = T$:

$$B = A(1 + \frac{R}{2})^{2T} \tag{7}$$

If $\underline{n = 12}$ (interest paid every month) at $t = T$:

$$B = A(1 + \frac{R}{12})^{12T} \tag{8}$$

If interest is paid n times a year then:

$$B = A(1 + \frac{R}{n})^{nT} \tag{9}$$

If the interest is paid continously, then n is a large number. In mathematics this translates to "$n \to \infty$". What we need to find out is what B is when $n \to \infty$, i.e., we want to find $lim_{n\to\infty}B = lim_{n\to\infty}A(1 + \frac{R}{n})^{nT}$.

We first assume that the limit exists and it is L:

$$L = lim_{n\to\infty} A(1 + \frac{R}{n})^{nT} \tag{10}$$

Taking the natural logarithm from both sides:

$$lnL = ln[lim_{n\to\infty} A(1 + \frac{R}{n})^{nT}] \tag{11}$$

$$lnL = lim_{n\to\infty} ln[A(1 + \frac{R}{n})^{nT}] \tag{12}$$

$$lnL = lim_{n\to\infty} ln[A] + ln[(1 + \frac{R}{n})^{nT}] \tag{13}$$

$$lnL = lim_{n\to\infty} ln[A] + (nT)ln[(1 + \frac{R}{n})] \tag{14}$$

Since n is a large number, $\frac{R}{n}$ will be a very small number close to zero. Hence $1 + \frac{R}{n}$ will be close to one. Using the taylor series of expansion and the rule that, $ln(1 + x) \to x$ for x close to zero:

$$lnL = lim_{n\to\infty} ln[A] + (nT)\frac{R}{n} \tag{15}$$

Canceling n:

$$lnL = lim_{n\to\infty} ln[A] + TR \tag{16}$$

There is no n left in the equation so we could get rid of "$lim_{n\to\infty}$". We are left with:

$$lnL = ln(A) + TR \tag{17}$$

Taking exponential from both sides:

$$lnL = exp(lnA + TR) = e^{lnA}e^{TR} = Ae^{TR} \tag{18}$$

Now that we have $B = Ae^{TR}$, we come back to what the rate of return is, $r_t = \frac{S_t - S_{t-1}}{S_{t-1}}$:

$$r_t = \frac{B_t - B_{t-1}}{B_{t-1}} = \frac{Ae^{TR} - Ae^{(T-1)R}}{Ae^{(T-1)R}} \tag{19}$$

$$r_t = \frac{Ae^{TR} - Ae^{(T-1)R}}{Ae^{(T-1)R}} = \frac{Ae^{R(T-1)}}{Ae^{R(T-1)}} \frac{e^R - 1}{1} \tag{20}$$

$$r_t = e^R - 1 \tag{21}$$

Hence:

$$e^R = r_t + 1 \tag{22}$$

Taking exponential from both sides:

$$R = ln(r_t + 1) \tag{23}$$

Replacing r_t with equation(1):

$$R = ln(\frac{P_t - P_{t-1}}{P_{t-1}} + 1) \tag{24}$$

$$R = ln(\frac{P_t}{P_{t-1}}) \tag{25}$$

$$R = ln(P_t) - ln(P_{t-1}) \tag{26}$$

Proof for the equation 3.38, spectral density of ARFIMA(0,d,0):

Recall that $x_t = (1 - L)^{-1}\epsilon_t = G(L)\epsilon_t$ for an ARFIMA (0,d,0) model. (Priestley, 1981) in section 4.12.3 on page 280 of his book states that if two stationary processes are linearly related then:

$$\sigma_x^2 f_x(\nu) = \sigma_\epsilon^2 |G(e^{-i\nu}|^2 f_\epsilon(\nu), \ -\pi \leq \nu \leq \pi \tag{27}$$

Hence:

$$\sigma_x^2 f_x(\nu) = W_x(\nu) = \sigma_\epsilon^2 |G(e^{-i\nu}|^2 f_\epsilon(\nu) = \sigma_\epsilon^2 |(1 - e^{i\nu})^{-d}|^2 f_\epsilon(\nu)$$

$$=\sigma_\epsilon^2 f_\epsilon(\nu)(1-e^{i\nu})^{-d}(1-e^{i\nu})^{-d}$$

$$=\sigma_\epsilon^2 f_\epsilon(\nu)(2-2cos(\nu))^{-d}, \quad as \quad e^{i\nu}=cos(\nu)+isin(\nu)$$

$$=\sigma_\epsilon^2 f_\epsilon(\nu)(2sin(\tfrac{\nu}{2}))^{-2d}, \quad as \quad 1-cos(\nu)=2sin^2(\tfrac{\nu}{2})$$

To prove that $lim\ W_x(\nu) \sim \sigma_\epsilon \nu^{-2d}$ as $\nu \longrightarrow 0$, we prove that:

$$lim_{\nu\longrightarrow 0}\ [\frac{2sin\frac{\nu}{2}}{\sigma_\epsilon^{\frac{-1}{2d}}\nu}]^{-2d}=1 \tag{28}$$

Which can be easily seen:

$$\lim_{\nu\longrightarrow 0}\ [\frac{2sin\frac{\nu}{2}}{\sigma_\epsilon^{\frac{-1}{2d}}\nu}]^{-2d}$$

$$=\lim_{\nu\longrightarrow 0}\ [\frac{2\frac{\nu}{2}}{\sigma_\epsilon^{\frac{-1}{2d}}\nu}]^{-2d}$$

$$=\lim_{\nu\longrightarrow 0}\ [\frac{\nu}{\sigma_\epsilon^{\frac{-1}{2d}}\nu}]^{-2d}$$

$$=\lim_{\nu\longrightarrow 0}\ [\frac{1}{\sigma_\epsilon^{\frac{-1}{2d}}}]^{-2d}$$

$$=\lim_{\nu\longrightarrow 0}\ =1$$

Proof is complete.

Driving equation 3.36 using spectral formulation:

There are two ways that one can get to equation 3.36, the autocovariance functions of ARFIMA(0,d,0). Either through the use of spectral formulation of the autocovariance functions or directly using the definition of the variable. We find autocovariance functions using both ways. First through the spectral formulation:

$$\kappa_\tau = \tfrac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\tau\nu} W_x(\nu) d\nu$$
$$= \tfrac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\tau\nu} \sigma_\epsilon^2 (2sin(\tfrac{\nu}{2}))^{-2d} d\nu$$
$$= (1/2\pi) \int_{-\pi}^{\pi} (cos(\nu\tau) + isin(\nu\tau)) \sigma_\epsilon^2 (2sin(\tfrac{\nu}{2}))^{-2d} d\nu$$
$$= (1/2\pi) \int_{-\pi}^{\pi} cos(\nu\tau) \sigma_\epsilon^2 (2sin(\tfrac{\nu}{2}))^{-2d} d\nu$$
$$= [\tfrac{\sigma_\epsilon^2}{2\pi}] \int_{0}^{2\pi} cos(\nu\tau) (2sin(\tfrac{\nu}{2}))^{-2d} d\nu$$

Using integration by substitution:

$$\kappa_\tau = [\frac{\sigma_\epsilon^2}{2\pi} 2^{-2d+1}] \int_0^{2\pi} cos(\frac{\nu}{2}2\tau)(sin(\frac{\nu}{2}))^{-2d} \frac{1}{2} d\nu$$

$$\kappa_\tau = [\frac{\sigma_\epsilon^2}{\pi}] 2^{-2d} \int_0^{2\pi} cos(t2\nu)(sin(t))^{-2d} dt, \;\; t = \frac{\nu}{2} \;\; dt = \frac{1}{2} d\nu$$

$$\kappa_\tau = [\frac{\sigma_\epsilon^2}{\pi}] 2^{-2d} [\frac{\pi cos(\pi\tau)\Gamma(-2d+2)2^{2d}}{(1-2d)\Gamma(\frac{-2d+1+2\tau+1}{2})\Gamma(\frac{-2d+1-2\nu+1}{2})}]$$

$$\kappa_\tau = \frac{\sigma_\epsilon^2 (-1)^k \Gamma(-2d+2)}{(1-2d)\Gamma(1+\tau-d)\Gamma(1-\tau-d)}$$

$$\kappa_\tau = \frac{\sigma_\epsilon^2 (-1)^k \gamma(1-2d)}{\Gamma(1+\tau-d)\Gamma(1-\tau-d)}, \;\; as \quad \Gamma(-2d+2) = (1-2d)\Gamma(1-2d)$$

$$\kappa_\tau = \frac{\Gamma(1-2d)\Gamma(\nu+d)}{\Gamma(d)\Gamma(1-d)\Gamma(\nu+1-d)} \sigma_\epsilon^2, \;\; as \quad \Gamma(\nu+d) = (-1)^k [\frac{\Gamma(d)\Gamma(1-d)}{\Gamma(1-\nu-d)}]$$

Gamma function has the asymptotic approximation that $\lim_{n\to\infty} \frac{\Gamma(n+\alpha)}{\Gamma(n)n^\alpha} = 1$, for $\alpha \in \mathbf{R}$ if $\alpha = 1-2d$, then:

$$\kappa_\tau = \frac{\Gamma(1-2d)\Gamma(\nu+d)}{\Gamma(d)\Gamma(1-d)\Gamma(\nu+1-d)} \sigma_\epsilon^2 \sim \frac{\Gamma(1-2d)}{\Gamma(d0\Gamma(1-d)} \nu^{2d-1} \sigma_\epsilon^2 \;\; as \;\; \tau \longrightarrow \infty \tag{29}$$

Driving equation 3.36 directly:

Recall from equation 3.67, that ARFIMA (0,d,0) can be obtained as:

$$x_t = (1-L)^{-d}\epsilon_t = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}\epsilon_{t-j}$$

So the autocovariance functions are:

$$\kappa_\tau = cov(x_t, x_{t-\tau})$$
$$\kappa_\tau = E(x_t \times x_{t-\tau}) - E(x_t)E(x_{t-\tau})$$
$$\kappa_\tau = E[(x_t) \times (x_{t-\tau})]$$
$$\kappa_\tau = E[(\sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}\epsilon_{t-j}) \times (\sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}\epsilon_{t-j-\tau})]$$
$$\kappa_\tau = E[(\epsilon_t + \frac{\Gamma(1+d)}{\Gamma(2)\Gamma(d)}\epsilon_{t-1} + \frac{\Gamma(2+d)}{\Gamma(3)\Gamma(d)}\epsilon_{t-2} + \ldots + \frac{\Gamma(\tau+d)}{\Gamma(\tau+1)\Gamma(d)}\epsilon_{t-\tau} + \frac{\Gamma(\tau+1+d)}{\Gamma(\tau+2)\Gamma(d)}\epsilon_{t-\tau-1}$$
$$+ \frac{\Gamma(\tau+2+d)}{\Gamma(\tau+3)\Gamma(d)}\epsilon_{t-\tau-2} + \ldots) \times$$
$$(\epsilon_{t-\tau} + \frac{\Gamma(1+d)}{\Gamma(2)\Gamma(d)}\epsilon_{t-\tau-1} + \frac{\Gamma(2+d)}{\Gamma(3)\Gamma(d)}\epsilon_{t-\tau-2} + \ldots)]$$

Since $E(\epsilon_{t-\tau} \times \epsilon_{t-\tau}) = E(\epsilon_{t-\tau-1} \times \epsilon_{t-\tau-1}) = E(\epsilon_{t-\tau-2} \times \epsilon_{t-\tau-2}) = \ldots$ we assume that the variance equals to σ_ϵ^2.

$$\kappa_\tau = \frac{\sigma_\epsilon^2}{\Gamma(d)^2} \times [\frac{\Gamma(\tau+d)\Gamma(d)}{\Gamma(\tau+1)} + \frac{\Gamma(1+d)\Gamma(\tau+1+d)}{\Gamma(2)\Gamma(\tau+2)} + \frac{\Gamma(2+d)\Gamma(\tau+2+d)}{\Gamma(3)\Gamma(\tau+3)} + \ldots]$$

$\Gamma(t)$ is defined as $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx$ and it follows from the definition that $\Gamma(t+1) = t\Gamma(t)$ and $\Gamma(n) = (n-1)!$. For example $\Gamma(\tau+1+d)$ can be written as $(\tau+d)\Gamma(\tau+d)$. The proof follows:

$$\kappa_\tau = \frac{\sigma_\epsilon^2}{\Gamma(d)^2}\frac{\Gamma(\tau+d)\Gamma(d)}{\Gamma(\tau+1)}[1 + \frac{(\tau+d)\times d}{(\tau+1)\times 1\times\Gamma(1)} + \frac{(\tau+d+1)(\tau+d)\times(1+d)(d)}{(\tau+2)(\tau+1)\times 2\times 1\times\Gamma(1)} +$$
$$\frac{(\tau+d+2)(\tau+d+1)(\tau+d)\times(2+d)(1+d)(d)}{(\tau+3)(\tau+2)(\tau+1)\times 3\times 2\times 1\times\Gamma(1)} + \ldots] =$$
$$\frac{\sigma_\epsilon^2}{\Gamma(d)^2}\frac{\Gamma(\tau+d)\Gamma(d)}{\Gamma(\tau+1)} \times F(\tau+d, d; \tau+1; 1)$$

A Gaussian hypergeometric function, is a series of the form as defined by (Gradshteyn and Ryzhik, 2007) (page 1005):

$$F(\alpha,\beta;\gamma;z)=1+\frac{\alpha\beta}{\gamma\times 1}z+\frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1)\times 1\times 2}z^2+\frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{\gamma(\gamma+1)(\gamma+2)\times 1\times 2\times 3}z^3+\dots \tag{30}$$

Which then satisfies this condition (Gradshteyn and Ryzhik, 2007) (page 1008):

$$F(\alpha,\beta;\gamma;z)=\frac{\Gamma(\gamma)\Gamma(\gamma-\alpha-\beta)}{\Gamma(\gamma-\alpha)\Gamma(\gamma-\beta)} \tag{31}$$

So κ_τ can be written as:

$$\kappa_\tau=\frac{\sigma_\epsilon^2}{\Gamma(d)^2}\frac{\Gamma(\tau+d)\Gamma(d)}{\Gamma(\tau+1)}\frac{\Gamma(\tau+1)\Gamma(1-2d)}{\Gamma(\nu+1-d)\Gamma(1-d)}=\frac{\sigma_\epsilon^2\Gamma(\tau+d)\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)\Gamma(\tau+1-d)}$$

Which is the same as equation 29.

From (Gradshteyn and Ryzhik, 2007) (page 895), Stirling's approximation is

$$\Gamma(z)\sim z^{z-\frac{1}{2}}e^{-z}\sqrt{2\pi} \tag{32}$$

and further $\lim_{|z|\to\infty}\frac{\Gamma(z+a)}{\Gamma(z)}e^{-alnz}=1$, which can be used to arrive at:

$$\lim_{|z|\to\infty}\frac{\Gamma(z+a)}{\Gamma(z+b)}=\frac{\frac{\Gamma(z+a)}{\Gamma(z)}}{\frac{\Gamma(z+b)}{\Gamma(z)}}=\frac{e^{alnz}}{e^{blnz}}=\frac{z^a}{z^b}=z^{a-b}$$

Proof is complete.

Driving equation 3.37, autocorrelation functions of ARFIMA(0,d,0):

$$\gamma_\tau=\frac{\kappa_\tau}{\kappa_0}=\frac{\frac{\Gamma(1-2d)\Gamma(\nu+d)}{\Gamma(d)\Gamma(1-d)\Gamma(\nu+1-d)}\sigma_\epsilon^2}{\frac{\Gamma(1-2d)\Gamma(d)}{\Gamma(d)\Gamma(1-d)\Gamma(1-d)}\sigma_\epsilon^2}$$

$$\gamma_\tau = \frac{\Gamma(d)\Gamma(1-d)\Gamma(1-d)\Gamma(1-2d)\Gamma(\nu+d)}{\Gamma(d)\Gamma(d)\Gamma(1-2d)\Gamma(1-d)\Gamma(\nu+1-d)}$$

$$\gamma_\tau = \frac{\Gamma(1-d)\Gamma(\nu+d)}{\Gamma(d)\Gamma(\nu+1-d)}$$

As $\lim_{n\to\infty} \frac{\Gamma(n+\alpha)}{\Gamma(n)n^\alpha} = 1, \qquad \alpha \in \mathbf{R}$, if replace $\alpha = 1-2d$, then:

$$\kappa_\tau = \gamma_\tau = \frac{\Gamma(1-d)\Gamma(\nu+d)}{\Gamma(d)\Gamma(\nu+1-d)} \sim \frac{\Gamma(1-d)}{\Gamma(d)}\tau^{2d-1} \ as \ \tau \longrightarrow \infty \tag{33}$$

Proof for equation 3.69, driving the autocovariance functions of temporally aggregated y_T:

We know from equations 3.63 and 3.67 that for an ARFIMA(0,d,0) time series (x_t) and its temporally aggregated series (y_T), we have $y_T = x_t + x_{t-1} + x_{t-2} + \ldots + x_{t-(h-1)}$ and $x_t = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}\epsilon_{t-j}$, where t represents units of fundamental time and $T = ht$ units of aggregate time. The autocovariance functions of an ARFIMA (0,d,0) given by equation 3.36 are:

$$\kappa_\tau^x = \frac{\sigma_\epsilon^2\Gamma(\tau+d)\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)\Gamma(\tau+1-d)}$$

We will also make use of the following equation:

$$\sum_{s=1}^{s=M} \frac{\Gamma(a+s)}{\Gamma(b+s)} = \frac{1}{1+a-b}\left[\frac{\Gamma(1+a+M)}{\Gamma(b+M)} - \frac{\Gamma(1+a)}{\Gamma(b)}\right] \tag{34}$$

Then the autocovariance function of y_T $(\kappa^y(\tau,h))$ is:

$$\kappa^y(\tau,h) = cov(y_T, y_{T-\tau})$$

$$\kappa^y(\tau,h) = cov(\textstyle\sum_{i=0}^{h-1} x_{t-i}, \sum_{i=h\tau}^{(h+1)\tau-1} x_{t-i})$$

$$\kappa^y(\tau,h) = cov(\textstyle\sum_{i=0}^{h-1} x_{t-i}, \sum_{i=0}^{h-1} x_{t-h\tau-i})$$

$$\kappa^y(\tau,h) = cov(x_t + x_{t-1} + \ldots + x_{t-(h-1)}, x_{t-h\tau} + x_{t-h\tau-1} + \ldots + x_{t-[(h+1)\tau-1]})$$

$$\kappa^y(\tau,h) = h\kappa^x_{h\tau} + (h-1)\kappa^x_{(h-1)\tau} + \ldots + \kappa^x_{h\tau-h+1} + (h-1)\kappa^x_{(h+1)\tau} + \ldots + \kappa^x_{h\tau+h-1}$$

$$\kappa^y(\tau,h) = -h\kappa^x_{h\tau} + \sum_{i=0}^{h-1}(h-i)\kappa^x_{(h\tau-i)} + \sum_{i=0}^{h-1}(h-i)\kappa^x_{(h\tau+i)}$$

$$\kappa^y(\tau,h) = -h\kappa^x_{h\tau} + \sum_{i=0}^{h-1}\sum_{k=1}^{h-i}\kappa^x_{(h\tau-i)} + \sum_{i=0}^{h-1}\sum_{k=1}^{h-i}\kappa^x_{(h\tau+i)}$$

$$\kappa^y(\tau,h) = \sum_{k=1}^{h}[-\kappa^x_{h\tau} + \sum_{i=0}^{h-k}\kappa^x_{(h\tau-i)} + \sum_{i=0}^{h-k}\kappa^x_{(h\tau+i)}]$$

Using equation 3.36, the proof follows:

$$\kappa^y(\tau,h) = \frac{\sigma^2\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}\sum_{k=1}^{h}[-\frac{\Gamma(h\tau+d)}{\Gamma(h\tau+1-d)}$$

$$+\sum_{i=0}^{h-k}\frac{\Gamma(h\tau-i+d)}{\Gamma(h\tau-i+1-d)} + \sum_{i=0}^{h-k}\frac{\Gamma(h\tau+i+d)}{\Gamma(h\tau+i+1-d)}]$$

$$\kappa^y(\tau,h) = \frac{\sigma^2\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}\sum_{k=1}^{h}[-\frac{\Gamma(h\tau+d)}{\Gamma(h\tau+1-d)}$$

$$+\sum_{i=1}^{h-k+1}\frac{\Gamma(h\tau-1+d-i)}{\Gamma(h\tau-d-i)} + \sum_{i=1}^{h-k+1}\frac{\Gamma(h\tau-1+d+i)}{\Gamma(h\tau-d+i)}]$$

Using equation 34:

$$\kappa^y(\tau,h) = \frac{\sigma^2\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}\sum_{k=1}^{h}[-\frac{\Gamma(h\tau+d)}{\Gamma(h\tau+1-d)} + \frac{1}{2d}[\frac{\Gamma(h\tau+d-h+k-1)}{\Gamma(h\tau-d-h+k-1)} - \frac{\Gamma(d+h\tau)}{\Gamma(h\tau-d)} + \frac{\Gamma(h\tau+d+h-k+1)}{\Gamma(h\tau-d+h-k+1)} - \frac{\Gamma(d+h\tau)}{\Gamma(h\tau-d)}]]$$

$$\kappa^y(\tau,h) = \frac{\sigma^2\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \times \sum_{k=1}^{h}[-\frac{\Gamma(h\tau+d)}{\Gamma(h\tau+1-d)} + \frac{2}{-2d}\frac{\Gamma(d+h\tau)}{\Gamma(h\tau-d)} + \frac{1}{2d}[\frac{\Gamma(h\tau+d-h+k-1)}{\Gamma(h\tau-d-h+k-1)} + \frac{\Gamma(h\tau+d+h-k+1)}{\Gamma(h\tau-d+h-k+1)}]$$

Using, $\Gamma(x+1) = x\Gamma(x)$ for $x = -d$:

$$\kappa^y(\tau,h) = \frac{\sigma^2\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \times \sum_{k=1}^{h}[-\frac{\Gamma(h\tau+d)}{\Gamma(h\tau+1-d)} + \frac{\Gamma(h\tau+d)}{\Gamma(h\tau+1-d)}] + \frac{1}{2d}[\frac{\Gamma(h\tau+d-h+k-1)}{\Gamma(h\tau-d-h+k-1)} + \frac{\Gamma(h\tau+d+h-k+1)}{\Gamma(h\tau-d+h-k+1)}]$$

$$\kappa^y(\tau,h) = \frac{\sigma^2\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \times \sum_{k=1}^{h}\frac{1}{2d}[\frac{\Gamma(h\tau+d-h+k-1)}{\Gamma(h\tau-d-h+k-1)} + \frac{\Gamma(h\tau+d+h-k+1)}{\Gamma(h\tau-d+h-k+1)}]$$

Replacing $h-k+1$ with j:

$$\kappa^y(\tau,h)=\frac{\sigma^2\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)2d}\times\sum_{j=1}^{h}[\frac{\Gamma(h\tau+d-j)}{\Gamma(h\tau-d-j)}+[\frac{\Gamma(h\tau+d+j)}{\Gamma(h\tau-d+j)}]$$

And using equation 34 again:

$$\kappa^y(\tau,h)=$$
$$\frac{\sigma^2\Gamma(1-2d)}{\Gamma(d+1)\Gamma(1-d)2(1+2d)}\times[\frac{\Gamma(1+h\tau+d-h)}{\Gamma(h\tau-d-h)}+\frac{\Gamma(1+h\tau+d+h)}{\Gamma(h\tau-d+h)}-2\frac{\Gamma(1+h\tau+d)}{\Gamma(h\tau-d)}]$$

Proof is complete.

Proof for equation 3.70, driving the autocorrelation functions of temporally aggregated y_T:

The autocorrelation functions for $y_T=y_{ht}$ are denoted as:

$$\gamma^y(\tau,h)=\frac{\kappa^y(\tau,h)}{\kappa^y(0,h)} \tag{35}$$

$$\kappa^y(0,h)=\frac{\sigma^2\Gamma(1-2d)}{\Gamma(d+1)\Gamma(1-d)2(1+2d)}\times[\frac{\Gamma(1+d-h)}{\Gamma(-d-h)}+\frac{\Gamma(1+d+h)}{\Gamma(-d+h)}-2\frac{\Gamma(1+d)}{\Gamma(-d)}]$$

Using the following equation from (Gradshteyn and Ryzhik, 2007) (page 896):

$$\Gamma(1-x)\Gamma(x)=\frac{\pi}{sin(\pi x)} \tag{36}$$

And replacing x with $-d+h$ and $-d-h$:

$$\frac{\Gamma(1+d-h)}{\Gamma(-d+h)}=\frac{\pi}{sin(\pi(-d+h))} \tag{37}$$

$$\frac{\Gamma(1+d+h)}{\Gamma(-d-h)}=\frac{\pi}{sin(\pi(-d-h))} \tag{38}$$

Since h is a positive integer and $sin(x)$ is a periodic function, $sin(\pi(-d+h))$ and $sin(\pi(-d+h))$ have the same value, hence:

$$\frac{\Gamma(1+d-h)}{\Gamma(-d-h)} = \frac{\Gamma(1+d+h)}{\Gamma(-d+h)} \tag{39}$$

And so:

$$\kappa^y(0,h) = \frac{\sigma^2\Gamma(1-2d)}{\Gamma(d+1)\Gamma(1-d)2(1+2d)} \times \left[2\frac{\Gamma(1+d-h)}{\Gamma(-d-h)} - 2\frac{\Gamma(1+d)}{\Gamma(-d)}\right]$$

Replacing above in equation 35:

$$\gamma^y(\tau,h) = \frac{\kappa^y(\tau,h)}{\kappa^y(0,h)}$$

$$\gamma^y(\tau,h) = \frac{\frac{\sigma^2\Gamma(1-2d)}{\Gamma(d+1)\Gamma(1-d)2(1+2d)} \times \left[\frac{\Gamma(1+h\tau+d-h)}{\Gamma(h\tau-d-h)} + \frac{\Gamma(1+h\tau+d+h)}{\Gamma(h\tau-d+h)} - 2\frac{\Gamma(1+h\tau+d)}{\Gamma(h\tau-d)}\right]}{\frac{\sigma^2\Gamma(1-2d)}{\Gamma(d+1)\Gamma(1-d)2(1+2d)} \times [2\frac{\Gamma(1+d-h)}{\Gamma(-d-h)} - 2\frac{\Gamma(1+d)}{\Gamma(-d)}]}$$

And hence:

$$\gamma^y(\tau,h) = \frac{\frac{\Gamma(1+h\tau+d-h)}{\Gamma(h\tau-d-h)} + \frac{\Gamma(1+h\tau+d+h)}{\Gamma(h\tau-d+h)} - 2\frac{\Gamma(1+h\tau+d)}{\Gamma(h\tau-d)}}{2\frac{\Gamma(1+d-h)}{\Gamma(-d-h)} - 2\frac{\Gamma(1+d)}{\Gamma(-d)}} \tag{40}$$

Proof is complete.

Proof for equation 3.71, the spectral density functions of aggregated y_T:

From equation 3.68, y_T takes the form:

$$y_T = \sum_{j=0}^{h-1}(1-L)^{-d}\epsilon_{t-j} = (1-L)^{(-d-1)}[1-L^h]\epsilon_t \tag{41}$$

Using equations 27 for y_T and using the same procedure, used for equation 3.38:

$$f^y(\nu,h) = \frac{\sigma^2}{2\pi}\left(2sin(\frac{\nu}{2h})\right)^{-2(d+1)}(2sin(\frac{\nu}{2})^2 \tag{42}$$

Proof is complete.

Proof for equation 3.73, the limiting $\gamma^y(\tau,h)$:

$\lim_{\tau\longrightarrow\infty}\gamma^y(\tau,h)$ can be donoted as:

$$\lim_{\tau\longrightarrow\infty}\frac{h^{2d+1}[(\tau+1)^{2d+1}+(\tau-1)^{2d+1}-2\tau^{2d+1}]}{2h^{2d+1}-2\dfrac{\Gamma(1+d)}{\Gamma(-d)}}$$

When $\tau\longrightarrow\infty$, the numerator is a $\infty+\infty-\infty$ case for which the common factor with the greatest exponent should be removed from the polynomial. Using, $(x+y)^n=\sum_{k=0}^{n}\binom{n}{k}x^{n-k}y^k$:

$$(\tau+1)^{2d+1} =$$
$$\binom{2d+1}{0}\tau^{2d+1}1^0+\binom{2d+1}{1}\tau^{2d}1^1+\binom{2d+1}{2}\tau^{2d-1}1^2+\binom{2d+1}{3}\tau^{2d-2}1^3+\ldots+\binom{2d+1}{2d+1}\tau^0 1^{2d+1} =$$
$$\tau^{2d+1}\left[1+(2d+1)\frac{1}{\tau}+d(2d+1)\frac{1}{\tau^2}+\frac{(2d+1)d(2d-1)}{3}\frac{1}{\tau^3}+\ldots+\frac{1}{\tau^{2d+1}}1^{2d+1}\right]$$

And

$$(\tau-1)^{2d+1} = \binom{2d+1}{0}\tau^{2d+1}(-1)^0+\binom{2d+1}{1}\tau^{2d}(-1)^1+\binom{2d+1}{2}\tau^{2d-1}(-1)^2+$$
$$\binom{2d+1}{3}\tau^{2d-2}(-1)^3+\ldots+\binom{2d+1}{2d+1}\tau^0(-1)^{2d+1} =$$
$$\tau^{2d+1}\left[1-(2d+1)\frac{1}{\tau}+d(2d+1)\frac{1}{\tau^2}-\frac{(2d+1)d(2d-1)}{3}\frac{1}{\tau^3}+\ldots+\frac{1}{\tau^{2d+1}}(-1)^{2d+1}\right]$$

And so $(\tau+1)^{2d+1}+(\tau-1)^{2d+1}-2\tau^{2d+1}$ can be written as:

$$(\tau+1)^{2d+1}+(\tau-1)^{2d+1}-2\tau^{2d+1}=$$
$$\tau^{2d+1}\left[1+(2d+1)\frac{1}{\tau}+d(2d+1)\frac{1}{\tau^2}+\frac{(2d+1)d(2d-1)}{3}\frac{1}{\tau^3}\right.$$

$$+\ldots+\tau^0 1^{2d+1}+1-(2d+1)\frac{1}{\tau}+d(2d+1)\frac{1}{\tau^2}-\frac{(2d+1)d(2d-1)}{3}\frac{1}{\tau^3}+\ldots+\tau^0(-1)^{2d+1}-2$$

$$=\tau^{2d+1}\left[2d(2d+1)\frac{1}{\tau^2}+\ldots\right]$$

$$=2\mathrm{d}(2\mathrm{d}+1)\tau^{2d-1}$$

And so:

$$\lim_{\tau\longrightarrow\infty}\frac{h^{2d+1}[(\tau+1)^{2d+1}+(\tau-1)^{2d+1}-2\tau^{2d+1}]}{2h^{2d+1}-2\dfrac{\Gamma(1+d)}{\Gamma(-d)}}$$
$$=\tau^{2d-1}\times\frac{h^{2d+1}\left[2d(2d+1)\right]}{h^{2d+1}-2\dfrac{\Gamma(1+d)}{\Gamma(-d)}}$$
$$=\tau^{2d-1}\times\frac{h^{2d+1}\left[d(2d+1)\right]}{h^{2d+1}-\dfrac{\Gamma(1+d)}{\Gamma(-d)}}$$

References

Ahmad, Y. and Paya, I. (2013). Temporal aggregation of random walk processes and implications for asset prices. Technical report, University of Wisconsin. 75

Amemiya, T. and Wu, R. Y. (1972). The effect of aggregation on prediction in the autoregressive model. *Journal of the American Statistical Association*, 67:628 – 632. 40, 74

Andersen, T. G. and Bollerslev, T. (1997a). Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *Journal of Finance*, 52(3):975 – 1005. 88, 107, 108, 109

Andersen, T. G. and Bollerslev, T. (1997b). Intraday periodicity and volatility persistence in financial markets. *Journal of empirical finance*, 4(2):115 – 158. 104, 107

Andersen, T. G. and Bollerslev, T. (1998a). Towards a unified framework for high and low frequency return volatility modeling. *Statistica Neerlandica*, 52:273 – 302. 87

Andersen, T. G. and Bollerslev, T. (1998b). Towards a unified framework for high and low frequency return volatility modeling. *Statistica Neerlandica*, 52(3):273 – 302. 87, 107, 109, 112

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001a). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61:43–76. 87, 94, 95

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2000). Great realizations. *Risk*, 13:105 – 108. 101, 104

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001b). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453). 75, 95

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001c). The distribution of realized exchange rate volatility. *NBER Working Paper*, 96:42 – 55. 92

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579 – 625. 96, 100

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2009). Parametric and nonparametric volatility measurement. *Handbook of financial econometrics*, 1:67 – 138. 95

Andersen, T. G., Bollerslev, T., and Meddahi, N. (2005). Correcting the errors: Volatility forecast evaluation using high-frequency data and realized volatilities. *Econometrica*, 73(1):279–296. 87, 96

Annunziata, M. (2011). *The Economics of the Financial Crisis*. Palgrave Macmillan. 2

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73:5 – 59. 54

Bandi, F. M. and Russell, J. R. (2008). Microstructure noise, realized variance, and optimal sampling. *The Review of Economic Studies*, 75(2):339 – 369. 101, 104

Barndorff-Nielsen, O. E. and Neil, S. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):253 – 280. 95

Baum, C., Barkoulas, J. T., and Caglayan, M. (1999). Long memory or structural breaks: can either explain nonstationary real exchange rates under the current float? *Boston College Working Papers in Economics*, 9(4):359 – 376. 57

Belsley, D. A. and Kontoghiorghes, E. J. (2009). *Handbook of computational econometrics*. Wiley Online Library. 11

Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman and Hall. 47, 50, 54, 67, 74

Bloomfield, P. (2000). *Fourier Analysis of Time Series: An Introduction*. John Wiley & Sons, New York. 10, 63

Bollerslev, T. and Wright, J. H. (2000). Semiparametric estimation of long-memory volatility dependencies: The role of high-frequency data. *Journal of Econometrics*, 98:81 – 106. 44, 45, 88, 112, 113

Box, G. E. P. and Jenkins, G. M. (1994). *Time Series Analysis: Forecasting and Control,3rd edition*. Pearson Education. 53

Box, G. E. P., Jenkins, G. M., and C., R. G. (2008). *Time Series Analysis: Forecasting and Control*. Wiley, Hoboken, NJ, 4th edition. 53

Burns, A. F. and Mitchell, W. C. . (1946). Measuring business cycles. *National Bureau of Economic Research*, pages 56 – 11. 10

Carmona, R. A., Hwang, W. L., and Torresani, B. (1998). *Practical TimeFrequency Analysis: Gabor and Wavelet Transforms with an Implementation in S*. Academic Press, San Diego. 19, 20, 21, 23

Chambers, M. J. (1998). Long memory and aggregation in macroeconomic time series. *International Economic Review*, pages 1053 – 1072. 40, 74, 81

Chan, N. H. (2010). *Time Series, Applicatins to Finance with R and S-plus*. John Wiley & Sons, Inc., second edition. 11

Chan, N. H. and Palma, W. (1998). State space modeling of long-memory process. *Annals of Statistics*, 26:719 – 740. 63

Christoffersen, P. F. and Diebold, F. X. (2000). How relevant is volatility forecasting for financial risk management? *Review of Economics and Statistics*, 82(1):12 – 22. 72

Christoffersen, P. F., Diebold, F. X., and Schuermann, T. (1998). Horizon problems and extreme events in financial risk management. *Economic Policy Review*, 4(3):98 – 16. 72

Chui, C. K. (1992). *An Introduction to Wavelets, Wavelet Analysis and its Applications*, volume 1. San Diego, Academic Press. 23

Coifman, R. R. and Donoho, D. L. (1995). Translation-invariant de-noising. *In Antoniadis and Oppenheim*, pages 125 – 150. 33

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223 – 236. 44, 45, 87, 91

Cont, R., Potters, M., and Bouchaud, J. P. (1997). Scaling in stock market data: stable laws and beyond. *Cornelll University Series*. 45, 91

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174 – 196. 54, 96

Crowley, P. M. (2007). A guide to wavelets for economists. *Journal of Economic Surveys*, 21(2):207 – 267. 23

Dacorogna, M., Gencay, R., Muller, U., Pictet, O., and Olsen, R. (2001). *An Introduction to High-Frequency Finance*. Academic Press. 65, 72, 88, 112

Dacorogna, M. M., Muller, U. A., Nagler, R. J., Olsen, R. B., and Pictet, O. V. (1993). A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *Journal of International Money and Finance*, 12(4):413 – 438. 92, 100, 104, 107, 109

Daubechies, I. (1992). *Ten Lectures on wavelets*. Volume 61 of CBMS-NSF Regional Conference Series in Applied Mathematics, Society of Industrial and Applied Mathematics, Philadelphia. 23, 27, 30, 31, 35

Diebold, F. X., Hickman, A., Inoue, A., and Schuermann, T. (1998). Scale models. *Risk 11*, pages 104 – 107. 72

Ding, Z., Granger, C. W. J., and Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1:83 – 106. 47

Doornik, J. A. and Ooms, M. (2003a). Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. *Computational Statistics and Data Analysis*, 42(3):333 – 348. 64

Doornik, J. A. and Ooms, M. (2003b). Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. *Computational Statistics & Data Analysis*, 42(3):333 – 348. 65

Drost, F. C. and Nijman, T. E. (1993). Temporal aggregation of garch processes. *Econometrica: Journal of the Econometric Society*, pages 909 – 927. 74, 112

Fan, J. and Wang, Y. (2007). Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association*, 102(480):1349 – 1362. 23

Fleming, J., Kirby, C., and Ostdiek, B. (2003). The economic value of volatility timing using realized volatility. *Journal of Financial Economics*, 67:473 – 509. 87

Forsberg, L. and Ghysels, E. (2007). Why do absolute returns predict volatility so well? *Journal of Financial Econometrics*, 5:31 – 67. 47

Fox, R. and Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series. *The Annals of Statistics*, 14:517 – 532. 57

Fraley, C., Leisch, F., Maechler, M., Reisen, V., and Lemonte, A. (2006). Fractionally differenced arima aka arfima (p, d, q) models. *R Package*, pages 1 – 3. 69

French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, 19(1):3 – 29. 94

Gallegati, M. (2008). Wavelet analysis of stock returns and aggregate economic activity. *Computational Statistics & Data Analysis*, 52(6):3061 – 3074. 23

Gao, R. X. and Yan, R. (2011). From fourier transform to wavelet transform: a historical perspective. In *Wavelets*, pages 17 – 32. Springer. 21

Gencay, R., Dacorogna, M. M., Olsen, R. B., and Pictet, O. V. (2003). Foreign exchange trading models and market behavior. *Journal of Economic Dynamics and Control*, 27(6):909 – 935. 36

Gencay, R., Selcuk, F., and Brandon, W. (2001a). Differentiating intraday seasonalities through wavelet multi-scaling. *Physica A: Statistical Mechanics and its Applications*, 289(3-4):543 – 556. 92, 108, 109

Gencay, R., Selcuk, F., and Whitcher, B. (2001b). Scaling properties of foreign exchange volatility. *Physica A: Statistical Mechanics and its Applications*, 289(1):249 – 266. 36, 72, 92

Gencay, R., Selcuk, F., and Whitcher, B. (2002). *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics.* Academic Press, USA, Texas. 23, 34, 38, 109

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4:221 – 238. 41, 57, 60

Golub, G. H. (1996). *Matrix Computations.* Johns Hopkins University Press, Baltimore, 3rd edition. 63, 64

Gonzaga, A. and Hauser, M. (2011). A wavelet whittle estimator of generalized long-memory stochastic volatility. *Statistical Methods & Applications*, 20(1):23–48. 83

Goodhart, C. A. E. and O Hara, M. (1997). High frequency data in financial markets: Issues and applications. *Journal of Empirical Finance*, 4(2):73 – 114. 88, 91, 104

Gradshteyn, I. S. and Ryzhik, I. M. (2007). *Table of Integrals, Series, and Products.* Academic Press, 7th edition. 126, 129

Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14(2):227 – 238. 54, 74

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1):15 – 29. 40, 53, 54, 55, 70, 74

Guillaume, D. M. (2000). On the intradaily performance of garch processes. In *Intradaily Exchange Rate Movements*, pages 95 – 115. Springer. 92

Hamilton, J. D. (1994). *Time Series Analysis.* Prinston University Press, New Jersey. 10

Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2):127 – 161. 100, 101, 104

Harvey, A. C. (1975). Spectral analysis in economics. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 24:1 – 36. 10

Harvey, A. C. (1991). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge. 10

Hassler, U. (2011). Estimation of fractional integration under temporal aggregation. *Journal of Econometrics*, 162(2):240 – 247. 40, 81

Heisenberg, W. (1930). *The Physical Principles of the Quantum Theory*. Dover Publication Inc., Chicago. 20

Hosking, J. R. M. (1981). Fractional differencing. *Biometrica*, 1(68):165 – 76. 40, 53, 54, 55, 57, 63

Hosking, J. R. M. (1996). Asymptotic distribution of the sample, mean, autocovariances and autocorrelations of long memory time series. *Journal of Econometrics*, 73:261 – 284. 63

Hubbard, B. B. (1998). *The World According to Wavelets, the story of a Mathematical Technique in the Making*. 2nd edition, A K Peters Ltd, USA. 23, 28

Hurst, H. E. (1951). Long-term storage capacity of resorvoirs. *Transactions of the American Society of Civil Engineers*, 116:277 – 302. 54, 57, 58

HWANG, S. (2000). The effects of systematic sampling and temporal aggregation on discrete time long memory processes and their finite sample properties. *Econometric Theory*, 16:347 – 372. 40, 74

Jacobsen, B. and Dannenburg, D. (2003). Volatility clustering in monthly stock returns. *Journal of Empirical Finance*, 10(4):479 – 503. 76

Kim, S. and In, F. (2010). Portfolio allocation and the investment horizon: a multiscaling approach. *Quantitative Finance*, 10:443 – 453. 36

Koopmans, L. H. (1995). *The Spectral Analysis of Time Series*. Academic Press, London. 10

Lo, A. (1991). Long-term memory in stock market prices. *Econometrica*, 59:1279 – 1313. 57, 58, 59, 62

Lo, A. and Mueller, M. (2010). Warning: physics envy may be hazardous to your wealth! *Journal of Investment Management*, 8:13 – 63. 72, 88

Lux, T. and Kaizoji, T. (2007). Forecasting volatility and volume in the tokyo stock market: Long memory, fractality and regime switching. *Journal of Economic Dynamics and Control*, 31(6):1808 – 1843. 54

Maheu, J. (2005). Can garch models capture long-range dependence? *Studies in Nonlinear Dynamics & Econometrics*, 9. 87

Malkiel, B. G. (2003). *A Random Walk Down Wall Street: The Time-Tested Strategy for Successful Investing*. WW Norton & Company, New York. 42

Malkiel, B. G. and Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work*. *The journal of Finance*, 25:383 – 417. 42

Mallat, S. (1998). *A Wavelet Tour of Signal Processing*. San Diego, Academic Press. 23, 25, 30, 31, 34, 35

Man, K. S. and Tiao, G. C. (2006). Aggregation effect and forecasting temporal aggregates of long memory processes. *International Journal of Forecasting*, 22(2):267 – 281. 40, 86

Mandelbrot, B. B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36(4):394 – 419. 40, 45, 98

Mandelbrot, B. B. (1971). When can price be arbitraged efficiently? a limit to the validity of the random walk and martingale models. *The Review of Economics and Statistics*, 53:225 – 236. 44

Mandelbrot, B. B. (1972). Statistical methodology for nonparametric cycles: From covariance to r/s analysis. *Annuals of Economic and Social Measurement*, 1:259 – 290. 57

Mantegna, R. N. and Stanley, H. E. (1995). Scaling behavior in the dynamics of an economic index. *Nature 376*, pages 46 – 49. 72

Marcellino, M. (1999). Some consequences of temporal aggregation in empirical analysis. 74

Martens, M. (2002). Measuring and forecasting s&p 500 index-futures volatility using high-frequency data. *Journal of Futures Markets*, 22(6):497–518. 76

Martens, M., Chang, Y. C., and Taylor, S. J. (2002). A comparison of seasonal adjustment methods when forecasting intraday volatility. *Journal of Financial Research*, 25(2):283 – 299. 109

McCoy, E. J. and Walden, A. T. (1996). Wavelet analysis and synthesis of stationary long-memory processes. *Journal of Comutational and Graphical Statistics*, 5:26 – 56. 32

Mcmillan, D. G. and Speight, A. E. H. (2008). Long-memory in high-frequency exchange rate volatility under temporal aggregation. *Quantitative Finance*, 8(3):251 – 261. 112, 113

Mihm, S. (2008). Dr. doom. *The New York Times Magazine*, page MM26. 2

Mincer, J. A. and Zarnowitz, V. (1969). The evaluation of economic forecasts. In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, pages 1 – 46. 114

Miron, J. A. (1996). *The Economics of Seasonal Cycles*. MIT Press, Massachusetts. 10

Nason, G. P. (2008). *Wavelet methods in statistics with R*. Springer. 23, 30, 33, 65, 109

Ohanissian, A., Russell, J. R., and Tsay, R. S. (2008). True or spurious long memory? a new test. *Journal of Business & Economic Statistics*, 26(2):161 – 175. 75

Olsen, R., Muller, U. A., Dacorogna, A. M., Olivier, P. V., Dave, R. D., and Dominique, G. M. (1997). From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and Stochastics*, 1(2):95 – 129. 104, 112

Osborn, D. R. and Ghysels, E. (2001). *The Econometric Analysis of Seasonal Time Series*. Cambridge University Press, Cambridge. 10

Pagan, A. (1996). The econometrics of financial markets. *Journal of Empirical Finance*, 3:15 – 102. 42

Palma, W. (2007). *Long-Memory Time Series: Theory and Methods*. Wiley, Hoboken, NJ. 53, 54, 63, 64

Pastor, L. and Veronesi, P. (2009). Technological revolutions and stock prices. *American Economic Review*, 99(4):1451 – 83. 40

Percival, D. B. and Guttorp, P. (1994). *Long memory processes, the Allan variance and wavelets, in Wavelets in Geophysics*. Academic Press, Sandiego, Calif. 33

Percival, D. B. and Walden, A. T. (2000). *Wavelet Methodology for Time Series Analysis*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press. 23, 29, 30, 32, 33, 34, 37, 38, 65, 69

Peter, C. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41(1):135 – 55. 72

Phillips, P. C. B. (1999). Discrete fourier transforms of fractional processes. Technical report, Cowles Foundation, Yale University. 61, 109

Pollock, D. S. G. (2012). *The Framework of a Dyadic Wavelet Analysis*. University of Leicester, Leicester. 35

Pong, S., Shackleton, M. B., Taylor, S. J., and Xu, X. (2004). Forecasting currency volatility: A comparison of implied volatilities and ar (fi) ma models. *Journal of Banking & Finance*, 28(10):2541 – 2563. 54

Power, G. J. and Turvey, C. G. (2010). Long-range dependence in the volatility of commodity futures prices: Wavelet-based evidence. *Physica A: Statistical Mechanics and its Applications*, 389(1):79 – 90. 23

Priestley, M. B. (1981). *Spectral Analysis and Time Series: Univariate Series*. Academic Press. 10, 122

Ramsey, J. B. (1999). The contribution of wavelets to the analysis of economic and financial data. *Philosophical Transactions of the Royal Society of London A*, 357:2593 – 2606. 8, 23, 65

Ramsey, J. B. (2002). Wavelets in economics and finance: Past and future. *Studies in Nonlinear Dynamics & Econometrics*, 6(3). 8, 21

Ramsey, J. B. and Lampart, C. (1998). The decomposition of economic relationships by time scale using wavelets: Expenditure and income. *Studies in Nonlinear Dynamics and Economics*, 3:23 – 42. 23

Ravishanker, N. (1997). Bayesian analysis of vector arfima processes. *Australian Journal of Statistics*, 39:295 – 312. 63

Reed, M. and Simon, B. (1976). *Methods of Modern Mathematical Physics II: Fourier Analysis and Self Adjointness*. Academic Press, New York. 20

Robinson, P. M. (1995). Log-periodogram regression of time series with long range dependence. *The annals of Statistics*, pages 1048 – 1072. 61

Robinson, P. M. (2003). *Time Series with Long Memory (Advanced texts in Econometrics)*. OUP Oxford, UK, Oxford. 53, 54

Roll, R. (1984). Orange juice and weather. *American Economic Review*, 74:861 – 880. 39

Rua, A. (2012). Wavelets in economics. *Economic Bulletin and Financial Stability Report Articles*. 8

Russell, J. R., Engle, R. F., Ait-Sahalia, Y., and Hansen, L. P. (2009). Analysis of high-frequency data. *Handbook of Financial Econometrics*, 1:383 – 426. 88

Schleicher, C. (2002). An introduction to wavelets for economists. Technical report. 23

Silvestrini, A. and Veredas, D. (2008). Temporal aggregation of univariate and multivariate time series models: A survey. Technical report, BANCA D'ITALIA. 74

Someda, C. G. (2006). *Electromagnetic Waves*. Taylor& Francis, second edition. 4

Souza, L. R. (2005). A note on chambers's long memory and aggregation in macroeconomic time series*. *International Economic Review*, 46(3):1059 – 1062. 40, 81

Souza, L. R. (2008). Spectral properties of temporally aggregated long memory processes. *Brazilian Journal of Probability and Statistics*, 22(2):135 – 155. 40

Souza, L. R. and Smith, J. (2002). Bias in the memory parameter for different sampling rates. *International Journal of Forecasting*, 18(2):299 – 313. 40, 74, 76

Souza, L. R. and Smith, J. (2004). Effects of temporal aggregation on estimates and forecasts of fractionally integrated processes: a monte-carlo study. *International Journal of Forecasting*, 20(3):487 – 502. 40, 81

Sowell, F. (1992a). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, 53:165 – 188. 57, 82

Sowell, F. (1992b). Modeling long-run behaviour with the fractional arima model. *Journal of Monetary Economics*, 53:165 – 188. 49, 63, 64, 65

Stanton, A. A., Day, M., and Welpe, I. M. (2010). *Neuroeconomics and the Firm*. Edward Elgar Publishing. 2

Stollnitz, E. J., Derose, T. D., and Salesin, D. H. (1996). *Wavelets for computer graphics: theory and applications*. Morgan Kaufmann Publishers, San Francisco. 8

Taleb, N. N. (2008). *The black swan: The impact of the highly improbable*. Penguin, London, England. 2

Taleb, N. N., Goldstein, D. G., and Spitznagel, M. W. (2009). The six mistakes executives make in risk management. *Harvard Business Review*, 87(10):78 – 81. 2

Taylor, J. W. (2010). Exponentially weighted methods for forecasting intraday time series with multiple seasonal cycles. *International Journal of Forecasting*, 26(4):627 – 646. 109

Tiao, G. C. (1972). Asymptotic behavior of temporal aggregates of time series. 59:525 – 531. 77

Ulrich, A. M., Dacorogna, M. M., Olsen, R. B., Pictet, O. V., Schwarz, M., and Morgenegg, C. (1990). Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis. *Journal of Banking and Finance*, 14(6):1189 – 1208. 89, 91, 100, 104, 109

Vacha, L. and Barunik, J. (2012). Co-movement of energy commodities revisited: Evidence from wavelet coherence analysis. *Energy Economics*, 34(1):241 – 247. 23

Vidakovic, B. (1999). *Statistical Modelling by Wavelets*. John Wiley Sons, New York. 29

Vuorenmaa, T. A. (2004). *A Multiresolution Analysis of Stock Market Volatility Using Wavelet Methodology*. PhD thesis, Department of Economics, University of Helsinki. 23

Wang, P. (2009). *Financial Econometrics*. Routledge, Abingdon, 2nd edition. 11

Whitcher, B. (2000). Maximum likelihood estimation for long memory time series using a non-decimated wavelet transform? *Geophysical Statistics Project, National Center for Atmospheric Research*. 41, 65, 67, 68, 69

Whittle, P. (1953). Estimation and information in stationary time series. *Arkiv for Matematik*, 2:423 – 434. 41, 65

Zaffaroni, P. (2007a). Aggregation and memory of models of changing volatility. *Journal of Econometrics*, 136:237 – 249. 74

Zaffaroni, P. (2007b). Contemporaneous aggregation of garch processes. *Journal of Time Series Analysis*, 28:521 – 544. 74

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005). A tale of two time scales. *Journal of the American Statistical Association*, 100(472). 102, 104

Zhang, Q. and Benveniste, A. (1992). Wavelet networks. *IEEE transactions on Neural Networks*, 3:889 – 898. 23

Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics*, 14(1):45 – 52. 100, 112